

No. 5

OCT 11 1949
September, 1949

Psychological Bulletin

EDITED BY

LYLE H. LANIER, NEW YORK UNIVERSITY

WITH THE CO-OPERATION OF

J. H. MITT, McCANN-BRICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HIRON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTH WESTERN UNIVERSITY; D. G. MARQUET, UNIVERSITY OF MICHIGAN; A. W. MELTZER, OHIO STATE UNIVERSITY; J. T. MITCALF, UNIVERSITY OF VERMONT.

CONTENTS

Reviews and Summaries:

Stimulus Generalization of Conditioned Responses: GREGORY RAZRAN, 337.

Psychotherapy as a Problem in Learning Theory: EDWARD J. SHOEN, JR., 366.

Statistical Methods Applied to Rorschach Scores: A Review: LEE J. CRONBACH, 393.

Reviews: 430.

Notes and Materials Received: 431.

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1215 Massachusetts Ave., N.W., Washington 5, D.C.

Subscription price, \$7.00 per year, single issue, \$1.25.

Second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879, authorized August 6, 1947.

Copyright, 1949, by The American Psychological Association, Inc.

PUBLICATIONS OF
The American Psychological Association, Inc.

AMERICAN PSYCHOLOGIST

Editor: DARL WOLFE, American Psychological Association
Contains all official papers of the Association and articles concerning psychology as a profession; monthly.
Subscription: \$7.00 (Foreign \$7.50). Single copies, \$.75.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

Editor: GORDON W. ALLPORT, Harvard University
Contains original contributions in the field of abnormal and social psychology, reviews, and case reports; quarterly.
Subscription: \$5.00 (Foreign \$5.25). Single copies, \$1.50.

JOURNAL OF APPLIED PSYCHOLOGY

Editor: DONALD G. PATTERSON, University of Minnesota
Contains material covering applications of psychology to business, industry, education, etc.; bi-monthly.
Subscription: \$6.00 (Foreign \$6.50). Single copies, \$1.25.

JOURNAL OF COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY

Editor: CALVIN F. STONE, Stanford University
Contains original contributions in the field of comparative and physiological psychology; bi-monthly.
Subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

JOURNAL OF CONSULTING PSYCHOLOGY

Editor: LAURANCE F. SHAFFER, Teachers College, Columbia University
Contains articles in the field of clinical and consulting psychology, counseling and guidance; bi-monthly.
Subscription: \$5.00 (Foreign \$5.50). Single copies, \$1.50.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Editor: FRANCIS W. IRWIN, University of Pennsylvania
Contains original contributions of an experimental character; bi-monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$1.25.

PSYCHOLOGICAL ABSTRACTS

Editor: C. M. LOUTIT, University of Illinois
Contains noncritical abstracts of the world's literature in psychology and related subjects; monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$.75.

PSYCHOLOGICAL BULLETIN

Editor: LYLE H. LANIER, New York University
Contains critical reviews of books and articles and critical and analytical summaries of psychological fields or subject matter; bi-monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$1.25.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Editor: HERBERT S. COMBES, U. S. Office of Education
Contains longer researches and laboratory studies which appear as units; published at irregular intervals at a cost to author of about \$2.50 a page; author receives 150 copies gratis.
Subscription: \$6.00 per volume of about 350 pages (Foreign \$6.50). Single copies, price varies according to size.

PSYCHOLOGICAL REVIEW

Editor: CARROLL C. PRATT, Princeton University
Contains original contributions of a theoretical nature; bi-monthly.
Subscription: \$5.50 (Foreign \$5.75). Single copies, \$1.50.

Subscriptions are payable in advance and are terminated at expiration.
Make checks payable to the American Psychological Association, Inc.

Subscriptions, orders, and other business communications should be sent to:

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVENUE N.W., WASHINGTON 9, D.C.

Psychological Bulletin

STIMULUS GENERALIZATION OF CONDITIONED RESPONSES¹

GREGORY RAZRAN

Queens College

INTRODUCTION

Historical. The doctrine of CR generalization and its presumed neurophysiological mechanism of irradiation and concentration entered Pavlov's system of behavior or, in his own words, "highest nervous activity," rather late. Extinction, the four kinds of internal inhibition (extinctive, differential, delayed, and conditioned), external inhibition (later discarded), inhibition of inhibition, and even analyzers, differential conditioning by contrasts, conditioning of higher order, and compound conditioning—all preceded it as concepts or phenomena. However, when irradiation-concentration was promulgated (34, 1910), it assumed a paramount position. Forming the core of all cortical dynamics, it became the "basic law of the highest² nervous activity" (34, p. 245) subsuming all older concepts and adding new ones, as subsidiaries, to account for alleged new CR discoveries. As is well known, Pavlov's entire system is conceptualized primarily upon terms borrowed, nominally, from classical neurophysiology. But one may also discern in his constructs the influence of theories of hydraulics and of sound transmission, the psychophysiology of G. H. Lewes, the logic and

¹ This study was prepared for publication while the writer was a Fellow of the John Simon Guggenheim Memorial Foundation.

² For convenience, the writer's references are to the English translations of the works of Pavlov and Bekhterev. The writer himself, however, uses the Russian originals, and his translations may thus differ in some way from those published. In this particular case, the Russian word VYSSHEY—a word that occurs also in the subtitle of the Pavlov book and quite commonly in the text—is translated by Gantt as *higher*. The writer used *highest* in abstracting the book later translated by Gantt, and then found that the German translators used *höchste*.

mental chemistry of J. S. Mill, and the stream of consciousness of William James.³

Pavlov's empirical findings of CR generalization in dogs were supposedly confirmed by Bekhterev, the neurologist and psychiatrist, in human adults (4), and by Krasnogorski, the pediatrician, in children (26). And both Bekhterev and Krasnogorski fully accepted Pavlov's neurophysiological interpretations (*op. cit.*). However, Beritoff, the Georgian (U.S.S.R.) physiologist, and Konorski, the Polish biologist, were highly critical of Pavlov's cortical constructs (5, 25), and it is known to the writer that a number of Russian psychologists were dubious even about some of his findings. Still, Pavlov's prestige was very high, and opposition to his views—except by Beritoff—was very much subvocal in nature. Criticisms came later, in the Thirties, but then they were not on experimental grounds. The teachings of Pavlov and Bekhterev were declared to be not accordant with Marxism: materialistic, yet asocio-political and insufficiently dialectical. The criticisms were directed mostly against the Human Reflexology of Bekhterev—even though Bekhterev himself wrote a long monograph (3) attempting to prove that Reflexology is Marxian in essence. Pavlov's laboratory, experimenting with animals, was given greater subsidies. Marxists, like Cartesians, draw sharp dichotomies between human beings and animals.

In this country, in the early days of Behaviorism, CR generalization was taken as a qualitative fact, and even demonstrated by Watson—who called it transfer—in his emotional conditioning of Albert (59). But when Pavlov's books came out in English (33, 1927; 34, 1928), a number of American students of behavior became critical of Pavlov's theories and skeptical of a good many of his facts. Lashley, himself a pioneer in CR experimentation, led in general criticism (28); Loucks, a young and able CR experimenter, took to task the doctrine of irradiation and indirectly that of generalization (30, 31); and Guthrie, an old-time Behaviorist, offered a new theory of generalization and of conditioning as such (11, 12, 13). On the other hand, Hull, quite affected by the quantitative and systematic implications of Pavlov's books, set up a special experiment (with M. J. Bass) (2) to test Loucks' scathing

³ From a few long conversations with Pavlov in the summer of 1934, the writer gained the impression that Pavlov was quite conversant with the writings of Wundt, James, and the British Associationists (Pavlov studied at, but did not graduate from, the Ryazan Theological Seminary). However, despite his gallant tribute to Thorndike (spelled Thorndyke in Anrep's translation—33, p. 6) as his predecessor in the objective analysis of behavior and his discussion of the views of Guthrie, Lashley, and Köhler in one of his articles (35), it is the writer's opinion that Pavlov had little familiarity with modern American comparative and systematic psychology.

analysis of CR irradiation and generalization. Hull was apparently satisfied that his results redeemed Pavlov's findings and accorded CR generalization systematic status as *postulate 5* in his system (20). Generalization, in general, became fashionable at Yale, in the mid-Thirties. Hovland started a series of experiments in the field (16, 17, 18); and Spence embarked upon a scheme of "deducing" discriminative learning from CR generalization (55, 56, 57). Hilgard and Marquis allotted an approving half-chapter to generalization in their book (15), and other textbooks followed in accepting it as an established fact, presumably an objective and quantitative substitute for the old "Law of Similarity" advocated by J. S. Mill and by Spencer but disputed by most other Associationists (8, 54, 58).

The Lashley-Hull controversy. The entire problem of CR generalization sprang up, however, again two years ago with a very stimulating article by Lashley and Wade (29). Lashley and Wade challenged everything in the doctrine: its facts, its interpretations, its significance. They cited Loucks' analysis, apparently not only for its specific criticisms, but also for its general implications that conclusions drawn by Pavlov and his students ought not to be taken at their face value. They stated that Bass and Hull (2) and Hovland (16, 17, 18) did not really corroborate primary stimulus generalization inasmuch as they used human subjects "for whom the stimulus series represented familiar relational sequences" and who may have used "habits of relational thinking" (p. 75). They mentioned the study by Wickens (60) who "failed to demonstrate a gradient of stimulus generalization" and one by the writer (48) who found "generalization with different types of stimuli too variable to formulate under any simple laws." Finally, Lashley and Wade detailed a series of their own experiments in all of which they failed to find generalization in animals or human subjects by their own technique of "training a group of subjects in a reaction to a single stimulus then opposing that stimulus to another on the same stimulus dimension and comparing the rates of formation of a discriminative habit when the reaction to the initial stimulus is reinforced and when it is extinguished by the differential training."

Lashley and Wade's interpretations are even more far-reaching than their criticisms of the experimental findings of CR generalization. These interpretations may be best summarized in their own words. They are:

1. There is no "irradiation" or spread of effects during primary conditioning (p. 74).
2. The phenomena of "stimulus generalization" represent a failure of association (p. 74) . . . failure to note the distinguishing characteristics of the stimulus or to associate them with the conditioned reaction (p. 81), . . . [and this generalization is thus really a] generalization by "default" (p. 82).

3. The "gradient of habit strength" is a product of variable stimulus thresholds, not a spread of associative process (p. 74). . . . The gradient varies with degree of attention and is unrelated to habit strength. . . . When inattention or threshold values are not involved, no evidence of a gradient is found. . . . Consequently, a test for "irradiation" may give the appearance of a gradient of habit strength when it is actually measuring discriminative thresholds under distraction (p. 84).

4. It [true generalization] does not occur in conditioning to a single stimulus but is somehow a function of differential training with two or more stimuli on the same dimension (p. 82). . . . The "dimension" of a stimulus series . . . do not exist for the organism until established by differential training (p. 74). . . . The dimension itself is created by or is a function of the organism and only secondarily, if at all, a property of the physically definable character of the stimuli (p. 82). . . . In the early stages of Pavlovian conditioning the only "dimension" common to such [tested for conditioning] stimuli is that all produce a sudden change in the environment. . . . With continued training the subject . . . may or may not show narrowing of the effective range on a stimulus dimension. Apparently such changes are a matter of chance noting of differences, generally with little regularity [reference to a study by the writer (48)] (p. 81).

Hull replied to Lashley and Wade (21). His reply is much more a defense of the experimental findings of CR generalization than a discussion of Lashley and Wade's special interpretations. He cites a private communication by Wickens that his [Wickens] data were in some respects quite harmonious with Hovland's findings and only when the Chi Square was calculated did "none of these differences [the differences between the conditioned and generalization stimuli] reach the 10 per cent level" (p. 130). He further states that in some of Hovland's experiments "the gradient at the first trial was horizontal" and that "If the gradient eventually found were due to the indirect effect of previously formed habits, e.g., of speech, it should have appeared at the very first trial" (p. 128). Hull then divulges unpublished results by Spence who found a stimulus generalization gradient with a Lashley and Wade technique, and reproduces seven graphs of CR generalizations from the studies of Anrep (1), Bass and Hull (2), Hovland (16, 17), Brown (7), and Wickens (60).⁴ Hull's own views need not be

⁴ Hull states: "The reference given by Lashley and Wade to Razran has been gone over with great care, but the present writer has been quite unable to find anything which would cast doubt on the reality of the falling gradient of stimulus generalization. . . . The article is essentially a contribution to a controversy concerned with response to explicit patterns of stimulations, a matter which is not under consideration here" (21, p. 130). Hull is obviously discussing another study by the writer (43) which is given in Lashley and Wade's bibliography, and not the particular one (48) to which they refer in the text. There is a typographical error in Lashley and Wade's reference: the writer's

gone into here since they are well known from his writings on this topic (19, 20, 21, 22). Hilgard (14, 1948) seems now to lean more to Lashley and Wade's views than to those of Pavlov and Hull in stating that "generalization may be along discriminated dimension" (p. 338). But a nearly straight line generalization curve with a modified Lashley and Wade technique was found by Grandine and Harlow (9) who, however, state that their findings "give no definitive answer as to whether this phenomenon of stimulus CR generalization is the result of some pre-established gradient around the training point, or is the specific resultant of the habit or habits established" (p. 336).

Background of the writer's interest. The writer has been interested in CR generalization since he published (with C. J. Warden) his first survey of pertinent Russian literature in 1929 (52). Through that survey and a number of others, the writer became not only critical of the theories of Pavlov and Bekhterev (39-42) but also convinced that the very empirical findings of the Russian laboratories need considerable checking and analysis. He came to learn that while the laboratories of Pavlov and Bekhterev are well advanced with respect to apparatus, physical set up of stimuli administration, controls of secondary cues, and refined measurements, their workers are very naive with regard to experimental design, sampling errors, and statistical treatments in general. Furthermore, it became known to him that individual CR experimenters in Pavlov's laboratory seldom exercise independence in the interpretation of their own data. Consequently, when the writer wished 10 years ago to review critically some significant aspect of conditioning, he chose extinction (45) rather than generalization, believing a good deal of the alleged phenomena of the latter to be little established. In 1937-1940 he carried out a series of experiments on the generalization of salivary CR's in adult human subjects. Three of the experiments have been published (44, 47, 48), but the publication of the rest has been delayed (three are in press now). In all three published studies, some CR generalization was manifested, but its course and very existence varied so much with the type of stimulus used that the writer stated in 1940 that his "empirical findings cast grave doubt upon attempts to 'deduce' transposition in discrimination experiments from generalization" and that laws of transfer "must be patiently discovered through experimentation and study of results at each level or form of stimulus, response, and organism organization" (48, p. 11).

mentioned study should have been referred to as "(19)" rather than "(18)." This study does not appear among the references in Hull's article.

Purpose of the present review. While the alleged empirical evidence for irradiation—the temporal after-effects of the application of conditioned stimuli—is, in the writer's opinion, worth checking and cannot be considered settled by Loucks' analysis,⁵ and while response generalization offers interesting possibilities,—there is no doubt that the important problem for students of behavior is that of stimulus generalization, or the more or less permanent CR influence of a conditioned stimulus upon a non-conditioned stimulus that is in some way related to it. And this is the problem which the present article purports to treat as fully as feasible. Specifically, the article will attempt, first, to examine all available evidence in answer to four empirical questions (three primary and one secondary) on CR generalization, and, then, to provide some integrative interpretations of the evidence and the answers.

The three primary empirical questions are:

1. Is there consistent evidence to show that an organism conditioned to some stimulus will also produce a CR in some degree to some other related stimulus, with which the organism has had no previous pertinent experience?
2. If the answer to the first question is positive, What is the relation of the strength of the CR to the related stimulus—magnitude, latency, resistance to extinction—to the strength of the CR to the conditioned stimulus?
3. Is there a gradient of CR generalization? And if there is, what does it correlate with? Is the correlation of such a nature as to permit some tentative mathematical treatment of the gradient?

The secondary empirical question is:

What is the relation between the amount of training that a CR has received and the answers to the first three primary questions? (Another secondary empirical question, the generalization of extinction cannot, unfortunately, be given consideration here, for lack of space.)

The evidence to be examined. The evidence for answering the empirical questions and for subsequent integrative interpretations of the answers will be sought primarily in (a) the studies from Pavlov's laboratory of salivary CRs in dogs (67 studies), (b) Yale studies of conditioning the GSR in adult human subjects (five studies), and (c) the writer's studies of salivary CRs in adult human subjects. Salivary conditioning of dogs constitutes probably more than 90 per cent of the CR work in the U.S.S.R., while there is reason for excluding from the present comparative analysis four other American studies, with data on

⁵ A preliminary analysis by the writer of a number of Russian studies, together with a closer re-examination—and some re-analysis—of Loucks' analysis, indicate that these after-effects certainly occur more often than chance would warrant. However, the writer agrees fully with Loucks that they are not of a nature to prove irradiation. They probably are results of some very initial differential conditioning.

CR experimentation. Of the four, the study by Wickens (60) is complicated by involving response as well as stimulus generalization and by the general vicissitudes of conditioning finger withdrawal. The experiment by Brown (7), besides being instrumental rather than classical conditioning, used intensities as generalization stimuli without being able to control the factor of intensity *per se* (see later sections of the present article). The two studies by the Lashley and Wade technique (9, 29) will be utilized but might be better kept apart with respect to direct comparisons. On the other hand, the Pavlov, the Yale, and the writer's studies have all used autonomic responses and classical techniques and have, in addition, the advantage that in the Yale studies the stimuli and in the writer's studies the response were the same as in Pavlov's studies, the major source of evidence.

THE EVIDENCE FROM PAVLOV'S LABORATORY

Source material. Prior to 1924, most—but not all—full reports of CR work in Pavlov's laboratory were published in the form of doctoral theses, in fulfillment of the requirements for the M.D. degree at the St. Petersburg Military Medical Academy where Pavlov served first as Professor of Pharmacology and then as Professor of Physiology. These theses are usually 100 to 200 pages long and as a rule represent, in the writer's estimate, about as much work as Ph.D. theses in physiology or psychology in a first-rate American university. Their tables, or rather protocols, are very detailed, trial-by-trial presentations of magnitudes and latencies of salivation, and of amounts and qualities of food seeking motor accompaniments, for each stimulus and each dog. The writer knows of 49 such CR theses from Pavlov's laboratory and he has read most of them (two of the six reports in Loucks' analysis are such theses).

Since 1924, however, the bulk of CR experimentation from Pavlov's laboratories has been published in the *Trudy Fiziologicheskikh Laboratorii Akademika I.P. Pavlova* (Transactions of the physiological laboratories of Academician I.P. Pavlov). The writer has in his possession ten volumes of these *Trudy*, from 1924 to 1941. These ten volumes contain 245 separate CR studies (nearly all salivation in dogs). The studies are as a rule less extensive than those in the earlier doctoral theses and, moreover, are not presented in great detail (the total number of pages in the ten volumes is a little over 3000). They were performed, not by candidates for the M.D. degree, but by Pavlov's numerous research assistants and associates, often physiologists of note in their own right.

Originally, the writer intended to analyze first the material of CR

generalization in the doctoral theses. But, unfortunately, these theses are not available to him this year, and he decided to confine himself to the data in the ten volumes of the *Trudy*. However, he does have impressions of the theses which he read prior to the preparation of this article.

Method of analysis. Only very few of the 245 CR studies in the *Trudy* deal specifically with CR generalization—or irradiation—which Pavlovians consider to have become established facts years ago. But a good number of the studies contain exact CR generalization data, incidental to the solution of some other problem in “cortical dynamics.” Very often a CR experimenter, after he has formed a CR to some stimulus, will try out the CR with some other non-conditioned stimulus or stimuli, and these other stimuli are quite commonly related in some ascertainable way to each other and to the conditioned stimulus. The writer has thus gone over a few thousand tables in the *Trudy*, and every time that he noted that a CR was tried with a non-conditioned stimulus, he computed the magnitude of its salivation as a per cent of the salivation to the conditioned stimulus, for the particular session and particular animal. In computing the percentages, account was of course taken of the equality of other experimental factors in the session; and to avoid the problem of differential conditioning, only the first two trials of a CR to a non-conditioned stimulus were used. Moreover, inasmuch as most experimenters from Pavlov’s laboratory present in each experiment the CR history of each dog, generalization stimuli that appeared in such histories were excluded from the computations.

As the data began accumulating, the writer decided to limit his task to four kinds of *dimensional* (other than intensity) CR generalizations, three kinds of *intensity* generalizations, and a number of what, for lack of another name, may be called *inter-dimensional* and *inter-sensory* CR generalizations. The four kinds of dimensional (other than intensity) generalizations included (a) frequencies of beats of metronomes, (b) frequencies of tones, (c) frequencies of rhythmic tactions, and (d) spatial distances of tactions. The three kinds of intensity generalizations contained intensities of flashes of lights, intensities of bells, and intensities of whistles. The inter-dimensional and inter-sensory CR generalizations refer to generalizations between such stimuli as metronomes and whistles, metronomes and lights or tactile stimuli, and the like.

More specifically, data were obtained for the following conditions:

1. All the 12 permutations of generalizations between metronomes, whistles, light flashes, and tactile stimuli;
2. Four dimensional (other than intensity) *positions*, with regard to fre-

quencies of metronomes, tones, and rhythmic tactions, and to spatial distances of tactions;

3. Three lower and three higher intensity positions, with respect to lights, whistles, and bells.

The dimensional and intensity positions were determined, with two exceptions, from the objectively given stimuli characteristics: frequencies, distances, c.p.s.'s, and decibels. The two exceptions were some cases in which spatial distances had to be estimated by the writer, and some cases in which the intensities of bells and whistles were not given in decibels (only the experimenters' mere descriptions of "weaker," "still weaker," "weakest," "stronger," "still stronger," and "strongest" being included). On the other hand, it should be pointed out that the magnitude of the distances between positions within dimensions and intensities were as a rule unequal, and that averaging had to be done by combining corresponding positions (i.e., I with I, II with II, III with III, and IV with IV).

The writer is fully aware of some of the pitfalls of his analysis. He would like to point out, however, first, that the Pavlov laboratories are extremely routinized with regard to experimenter, design, apparatus, stimuli, and responses; and, second, that extreme care was taken by him to see that each of the nearly 700 comparisons between the CR's to the generalization and to the conditioned stimuli was performed under wholly equal conditions. Moreover, while a combined analysis may have masked the effects of some variables, one of these variables, the amount of training that the CR has received, has been singled out for special study. The data from frequencies of metronomes, spatial distances of tactions, intensities of lights, and transfers from metronomes to lights have been fractionated so as to compare CR generalization after 1-20, 21-40, 41-100, 101-300, and more than 300 reinforcements. Similar treatments of other variables—which the Russians maintain affect CR generalization, such as the age of the animals, cortical extirpation, certain drugs, and a few more—would have been possible but were not considered worth while undertaking, at least for the time being. It is believed, though, that the alleged effects of these variables might well have been cancelled out, since the writer included in his combined analysis data from both supposed generalization-increasing and supposed generalization-decreasing studies.

Results for Dimensional (other than intensity) CR Generalization

All the *Trudy* data on dimensional (other than intensity) CR generalization are combined in Table I. Each entry in the table is a

mean per cent of generalization of a number of determinations (exact number is given in parentheses), obtained from 8 to 67 different experiments. The *t*'s and *P*'s of the entries in Table I—as well as of the entries in the subsequent tables—have all been ascertained, but their separate tabulation would unduly lengthen the article. The most pertinent ones will be taken up in the text, in connection with each finding.

TABLE I

DIMENSIONAL (OTHER THAN INTENSITY) STIMULUS GENERALIZATION OF SALIVARY
CONDITIONING IN DOGS. DATA FROM 67 DIFFERENT EXPERIMENTS
IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Conditioned Stimuli	Steps Removed from Conditioned Stimuli			
	I	II	III	IV
Metronome	61(28)	57(22)	43(16)	45(11)
Tone	69(23)	59(19)	46(13)	42(10)
Spatial Taction	78(18)	59(14)	63(9)	54(8)
Rhythmic Taction	51(16)	57(13)	41(8)	44(8)

As seen from Table I, the answers to the first two of the previously posed "primary empirical questions" (*supra*, p. 342) are unmistakable. Non-conditioned generalization stimuli do evoke conditioned responses, and these responses are smaller in magnitude than those evoked by the conditioned stimuli. All the *t*'s between the magnitudes of the conditioned and the generalization CR's are significant with *P* equalling .01, the smallest *t* being 17.4. However, the answer to the third question, the question of a CR gradient, is by no means certain. Four of the 12 differences between adjacent generalization stimuli are reversals, and four of the remaining eight differences are insignificant, with *P* equalling .05. Yet *there would be no reversals and differences between adjacent generalization stimuli would be significant if the four dimensional steps were reduced to two by combining the first step with the second and the third with the fourth.* The importance of the last statement, as well as the fact that two of the four reversals occurred between the third and the fourth dimensional steps, will be discussed in a later section.

Results for Intensity CR Generalization

All the data on CR generalization of intensities are contained in Table II. They are given in percentages, and separately for each of the three lower and each of the three higher intensities. Unlike Table I, the gradients in Table II are quite consistent. There are no reversals, and

all 12 differences between adjacent stimuli are significant with P equaling .05. However, Table II shows *two* gradients: a descending one for lower intensities but an ascending one for higher intensities. Indeed, it is very doubtful whether intensity generalizations could be at all con-

TABLE II

STIMULUS INTENSITY GENERALIZATION OF SALIVARY CONDITIONING IN DOGS.
DATA FROM 54 DIFFERENT EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Conditioned Stimuli	Lower Intensity Steps			Higher Intensity Steps		
	I	II	III	I	II	III
Lights	79(14)	69(11)	58(8)	118(14)	128(9)	149(8)
Whistles	68(12)	58(11)	49(9)	137(13)	149(8)	165(8)
Bells	72(11)	64(9)	56(7)	124(10)	138(8)	149(6)

sidered true CR generalizations (cf. Hovland, 17). At any rate, these generalizations are certainly *sui generis*, and little ought to be inferred from them to generalizations within non-intensity dimensions. Furthermore, there is reason to believe that an intensity-like factor, namely, psychological intensity resulting from summation of stimuli, masks true generalizations along such stimuli dimensions as frequencies of metronomes and of tactile vibrators.

Results for Inter-Dimensional and Inter-Sensory CR Generalizations

What is striking about this type of CR generalization, as disclosed in Table III, is its large amount. There is, for instance, more CR generalization from a metronome to a whistle than from one metronome to another three steps removed, and there is nearly as much generalization from a light to a whistle than from one tone to another four steps away.⁶ These results, among others, cast grave doubt, in the writer's opinion, upon any postulate that the magnitude of CR generalization merely varies with the degree of relatedness of the generalization stimuli to the conditioned stimuli in some stimulus dimension or dimensions. A better case could perhaps be made out for a variation with some denotative relatedness or relatednesses in a phenomenal world.

⁶ It should be mentioned, however, that it is very much easier to establish differential conditioning between inter-dimensional and inter-sensory stimuli than between stimuli on the same dimension.

*Results for CR Generalization as a Function of the Amount
of Training of the CR*

With the exception of intensity generalization which was already noted as a special case, all the data in Table IV point to the following conclusion: (a) that CR generalization increases in the very initial stages of training the CR; (b) with further training of the CR, it begins

TABLE III

INTER-DIMENSIONAL AND INTER-SENSORY STIMULUS GENERALIZATION OF SALIVARY
CONDITIONING IN DOGS. DATA FROM 58 DIFFERENT EXPERIMENTS
IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

<i>Conditioned Stimulus</i>	<i>Generalization Stimulus</i>	<i>Per Cent of Generalization</i>
Metronome	Whistle	44(33)
Metronome	Light	38(31)
Metronome	Rhythmic taction	40(27)
Whistle	Metronome	42(21)
Whistle	Light	36(19)
Whistle	Rhythmic taction	38(20)
Light	Metronome	38(22)
Light	Whistle	39(24)
Light	Rhythmic taction	34(18)
Rhythmic taction	Metronome	41(19)
Rhythmic taction	Whistle	36(21)
Rhythmic taction	Light	32(17)

to decrease slowly; and (c) after a great number of reinforcements, the generalization may increase again. All the *t*'s between 1-20 and 21-40 reinforcements are significant (with *P* equalling .01), while the *t*'s between 21-40 and 41-100 and those between 101-300 and more than 300 reinforcements are significant (with *P* equalling .05), for frequencies of metronomes and for spatial distances.

Summary of the Evidence from Pavlov's Laboratory

A statistical analysis of the data on CR generalization contained in 67 experiments, performed between 1924 and 1941, in Pavlov's laboratory provides full and clear answers to the first, second, and fourth questions posed in the first section of this article but offers only a partial

TABLE IV

STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN DOGS AS DEPENDENT UPON THE NUMBER OF REINFORCEMENTS OF THE CONDITIONED STIMULUS.

DATA FROM 67 EXPERIMENTS IN PAVLOV'S LABORATORY

Each entry is a mean per cent of conditioned salivation to the non-conditioned generalization stimuli. Figures in parentheses are numbers of determinations.

Types of CR Generalizations	Number of Reinforcements of the Conditioned Stimulus				
	1-20	21-40	41-100	101-300	300 up
Frequencies of Metronomes	39(12)	64(17)	55(16)	49(12)	58(15)
Spatial Distances of Taction	54(8)	72(10)	64(9)	61(7)	74(11)
Metronomes to Lights	31(6)	42(8)	41(5)	36(5)	39(7)
Intensities of Lights*	96(10)	99(11)	103(14)	100(12)	96(17)

* Combining lower and higher intensities.

and not wholly certain answer to the third question. Specifically, the answers are:

1. Dogs conditioned to secrete saliva upon the application of some conditioned stimulus will also produce conditioned salivation upon the application of related stimuli with which the animals have had no previous pertinent experience.

2. Unless the intensity of the related stimulus is considerably higher than that of the conditioned stimulus, the magnitude of the related generalization CR will be considerably smaller than that of the trained CR.

3. CR generalization increases in the very initial stages of training the CR but upon further training begins to decrease slowly, while after a large number of reinforcements (over-training) it may increase again.

4. The gradient of CR generalization is very crude, consisting only of two or three steps.

5. The gradient does not vary *merely* with the degree of relatedness of the generalization stimuli to the conditioned stimuli along some stimulus dimension or dimensions.

THE YALE STUDIES

All the Yale studies to be analyzed here have been performed with human subjects and with the GSR as the conditioned response. Unlike the studies from Pavlov's laboratory, the Yale investigations have mastered, as might have been expected, not only the physical setups but also the design and the statistical treatments of their problems. Indeed, Hovland's three experiments may well be held up as CR Ph.D. paradigms. Moreover, there is little doubt that individual experimenters at Yale exercise much more independence in interpretations than individual experimenters in Pavlov's laboratory.

Hovland's Experiment with Frequencies of Tone (16)

This experiment has been quoted so often as an exemplar of a lawful and consistent CR gradient that the writer takes the liberty to reproduce here, for purposes of analysis, its main, and practically only, table (there are two more single line tables in the article). Hovland's table becomes Table V in this article, and the reader is asked to examine it very carefully.

TABLE V*

GENERALIZATION OF EXCITATORY TENDENCIES

Amplitudes of galvanic responses (in mm.) to conditioned tone (0) and to tones 25(1), 50(2), and 75(3) j.n.d.'s removed in frequency. Each value is average of two determinations following 16 reinforcements. Subjects I-X were conditioned to tone of 153 cycles; subjects XI-XX to tone of 1967 cycles.

Subject	Tonal Stimuli			
	0	1	2	3
I	16.2	11.3	12.6	11.4
II	22.4	18.1	25.4	20.9
III	13.5	6.7	11.2	6.3
IV	15.3	6.4	3.3	7.6
V	19.2	18.5	22.8	15.3
VI	16.2	18.5	11.9	13.9
VII	23.7	17.1	18.0	16.5
VIII	11.5	14.1	9.6	13.8
IX	13.8	10.3	13.4	9.9
X	22.4	18.7	10.7	15.3
XI	23.7	21.3	21.1	10.2
XII	16.5	20.6	13.9	12.4
XIII	17.7	18.4	15.2	13.7
XIV	18.6	13.9	14.3	12.5
XV	15.9	15.5	13.8	14.2
XVI	18.8	12.3	14.6	9.7
XVII	21.3	9.7	10.5	12.3
XVIII	23.2	17.8	13.9	14.5
XIX	13.8	16.8	9.3	11.8
XX	19.9	12.3	6.9	15.6
Mean	18.3	14.91	13.62	12.89
P.E.M.	0.57	0.64	0.8	0.49

* Reproduced from Hovland with permission.

As seen from Table V, the mean magnitude of the CR to the conditioned stimulus was 18.3 ± 0.57 (P.E.) mm., while the mean magnitudes of the generalization CR's—removed 25, 50, and 75 j.n.d.'s from the conditioned stimulus—were respectively: 14.91 ± 0.64 , 13.6 ± 0.80 , and

12.89 \pm 0.49 mm. The differences between the magnitude of the conditioned CR and the generalization CR's are fully reliable statistically, and so is the difference between the CR to the first and to the third generalization stimulus (Diff./P.E._{diff} = 4.21). *But the differences between adjacent generalization stimuli, i.e., between the first and the second and between the second and the third are not reliable statistically and are rather small.* Furthermore, if we examine the records of individual subjects, we find the following:

1. Nineteen of the 20 subjects showed one or more reversals, i.e., had mean CR's to more remote stimuli greater than to less remote stimuli;

2. Seven of the 19 subjects had mean CR's to generalization stimuli greater than the mean CR's to the conditioned stimulus;

3. The 20th subject, who showed no reversals, hardly manifested any gradient between the second and the third generalization stimuli—respective CR's being 21.3 and 21.1 mm.

Thus, not a single of Hovland's 20 subjects revealed a consistent gradient of CR generalization and only one of the three gradient-determining differences was statistically reliable.

Hovland's Study with Varying Intensities of Tones (17)

Two equated groups of 16 subjects each were used. The conditioned stimuli were a tone of 86 decibels for one group and a tone of 40 decibels for the other group, while the generalization stimuli were three descending tones 50, 100, and 150 j.n.d.'s removed from the conditioned stimulus for the first group, and three similarly removed ascending tones for the second group. Each conditioned stimulus and each generalization stimulus was tested three times, after 16 reinforcements of the conditioned stimulus with the unconditioned stimulus of electric shock. The mean CR to the conditioned tone of 86 decibels was 17.85 mm. while the respective CR's to the descending generalization tones were: 13.37, 10.94, and 8.9 mm. But the mean CR to the conditioned tone of 40 decibels was 10.3 mm. while the respective generalization CR's were: 12.98, 14.3, and 15.75 mm. Combining the results of the two groups, Hovland obtained a mean CR of 14.3 \pm 0.6 to the conditioned tones and means of 13.7 \pm 0.57, 13.17 \pm 0.57, and 12.62 \pm 0.44 to the three non-conditioned generalization tones. Hovland states that such combining "enables one to determine the gradient of *intensity* generalization with the intensity effect *per se* held constant" (p. 282) and believes that it is justifiable "because the relationship between intensity and magnitude of response conditioned is linear" (p. 285). This linearity cannot be considered wholly established. But even, if it were established, this gradient of intensity generalization with intensity *per se* held constant is certainly, as Hovland's data indicate, very small and very unreliable statistically.

*Hovland's Study of CR Generalization as a Function of
Amount of Reinforcement (18)*

Hovland's findings here are so much like those from Pavlov's laboratory—namely, an increase in generalization in early stages of training the CR, then a slow decrease and then again an increase—that they need not be restated. The only difference between the types of studies is that the decrease in generalization began in Hovland's subjects after 16 reinforcements and in Pavlov's dogs after about 40 reinforcements, a difference that is only apparent, since more reinforcements are as a rule required to form a salivary CR in a dog than a conditioned GSR in a man.

Humphreys' Experiment (23)

Humphreys used two methods of reinforcement, a regular and a special, in his experiment with conditioning frequencies of tones. With the regular method, used for 34 subjects, the conditioned stimulus was always reinforced; while with the special method, used for 20 subjects, only half of the applications of the conditioned stimulus were accompanied by reinforcement (electric shock). One of Humphreys' three generalization stimuli was 5 j.n.d.'s removed from the conditioned stimulus (a tone of 1967 cycles), and another was 15 j.n.d.'s removed. The third generalization stimulus, on the other hand, was removed 25 j.n.d.'s from the conditioned stimulus in one half of the subjects. But in the other half it was a tone 26 j.n.d.'s removed, but the lower octave of the conditioned stimulus. Humphreys' results show the mean magnitude of the CR to the conditioned tone to have been 3.52 mm., with magnitudes of 2.91 for the tone 5 j.n.d.'s removed, 2.84 for the 15 j.n.d. tone, 2.45 for the 25 j.n.d., and 3.27 mm. for the lower octave—when the regular method of reinforcement was used. With the special method of 50 per cent reinforcement, the mean magnitude of the CR to the conditioned tone was 3.94 mm., while the mean magnitudes of the generalization CR's were respectively: 3.76, 4.19, and 3.13 mm. (The last figure is a combination of the CR's to the tone 25 j.n.d.'s removed and to the octave 26 j.n.d.'s away, no separate figures being given by the experimenter.) Thus, we have here an experiment on CR generalization that shows a reversal and very small and unreliable differences between CR's to generalization stimuli differing widely in dimensional relatedness. Hull (21) does not cite Humphreys' study in his marshalling of evidence for a true gradient of CR stimulus generalization. But Lashley and Wade (29) do mention it.

The Bass and Hull Experiment (2)

Bass and Hull worked with spatial generalization of taction, using eight subjects. The generalization stimuli were spaced approximately

16, 32, and 48 inches from the conditioned spot, and each generalization stimulus was tested 32 times. The mean CR to the conditioned stimulus was 5.74 ± 0.39 mm., while the means to the generalization stimuli were respectively: 5.63 ± 0.49 , 4.75 ± 0.395 , and 3.36 ± 0.32 mm. Four of the eight subjects showed reversals, and the subject for whom data are presented on four successive days showed reversals on two of the four days. In a certain way, however, the gradient obtained by Bass and Hull is the sharpest from the Yale laboratories. But, of course, it too—with reversals in half of the subjects and practically no difference between the group CR to the conditioned stimulus and to the generalization stimulus 16 inches away—could hardly support the conclusion by Hull that "This generalization gradient of reaction strength is a monotonic decreasing function of the magnitude of the differences between the conditioned and the unconditioned⁷ [generalization] stimuli" (21, p. 133). Furthermore, there is a serious criticism of the Bass and Hull study, as far as the gradient is concerned, namely: their alternate reinforcement of the conditioned stimulus and non-reinforcement of the generalization stimuli permitted differential conditioning by contrasts to develop.

Summary and Evaluation of the Yale Studies

There is on the whole a striking similarity between the results on CR generalization in the Yale and in the Pavlov studies. Both provide about the same answers to the four questions asked in the first section of this article. The fact that the Yale studies were performed with human subjects makes their gradient results subject to interpretations of "habits of relational thinking" (Lashley, 29) and of earlier experience with "conventional dimensions of discriminations" (Hilgard, 14). Hull's statement about the "relatively non-voluntary character of galvanic skin reactions" (21) falls short in that in generalization it is the consciousness of the stimuli, not only of the responses, that counts, and human subjects are certainly conscious of the stimuli in GSR experiments, and consciousness of stimuli certainly affects the GSR.

However, while human subjects could use the devices mentioned by Lashley and by Hilgard, they of course need not use them, and the very crude and irregular gradients obtained could be construed as evidence that they did not, at least consistently, use them. But this may be answered by the argument that the "sets" in the Yale studies did not call for conscious and active discriminations, so that these have manifested themselves only occasionally and incidentally (something like "incidental memory" or "premature reactions" in reaction time experi-

⁷ *Unconditioned* is a very confusing adjective here. *Non-conditioned* would be better.

ments). On the other hand, the close similarity between the Yale and the Pavlov results should lend support to the supposition that previous experience was not much operative in the Yale studies. Ultimately the problem seems to resolve itself into the alternative of the Yale findings being results of (a) relational thinking superimposed upon a true CR gradient or (b) relational thinking superimposed upon chance, with the effects of relational thinking varying possibly from zero to 100 per cent.

RAZRAN'S EXPERIMENTS

All experiments on CR generalization by the writer were with salivary CR's in college undergraduates. His general technique of salivary conditioning has been described previously (42, 44, 46) and need not be repeated here. All that might be added are three of its more recent characteristics; viz.: (a) multiple intermittent one-second presentations of stimuli-to-be-conditioned during single continuous eating periods of two to four minutes, so as to provide maximum attention for the stimuli and to make the entire task more "molar" and meaningful; (b) misinforming the subjects about the nature of the experiment, so as to forestall disrupting subjective attitudes; and (c) varying the food and scheduling the sessions in late mornings and afternoons, so as to insure adequate psychophysiological motivation. Of the three characteristics, the most important one is, in a way, "misinforming the subjects" such as telling them that the experiment aims to study "the effects of hunger or satiety or digestion upon eye-fatigue" (with visual stimuli) or "upon ear-fatigue" (with auditory stimuli) or "upon memory" (with verbal stimuli), and actually administering some "sham" tests of fatigue or memory. Thus, the writer's subjects are very attentive to the stimuli, the responses, their own physiological and mental states; but wholly unaware of the experimenter's objective and his attempt to condition them. Their "consciousness" and "sets" have been "naturally" diverted. (Occasionally a subject will "catch on," and then his results are treated separately.)

Experiments with Sensory Stimuli

Two types of experiments were performed. In one experiment, used with 32 subjects, the conditioned-stimuli were the flashes of two to ten miniature spherical lights ($\frac{3}{8}$ inches in diameter, 2 c.p., 2.5 volts) of the same or of different colors. The lights were arranged in different spatial patterns—many of the patterns being the same as the dot arrangements of Schumann, Rubin, and Wertheimer—and they were flashed either simultaneously or in specially selected temporal sequences. Their generalization stimuli were (a) fewer lights, (b) lights

differing in color, (c) lights in different spatial arrangements, and (d) lights in different temporal sequences. In the other experiment, used with four subjects, the conditioned stimuli were 12 musical intervals (excluding the natural minor seventh) in the c^2 - c^3 octave of a Stoelting *just* harmonium. The generalization stimuli were different musical intervals in the same octave and with one common tone, different musical intervals in the same octave but with no common tone, the same musical intervals in a different octave and in a different key, and the same musical intervals in a different octave but in the same key.

There was no attempt in these experiments to plot CR gradients but to a) establish the existence of CR generalization, in general; b) study its dependence upon amount of reinforcement of the CR; and particularly c) study the dependence of CR generalization upon (1) the spatial, temporal, and color patterns in the case of lights, and upon (2) common ratios, partials, and absolute tones in the case of the musical intervals. The results showed: (a) unmistakable evidence for the existence of CR generalization and for the lesser magnitudes of generalization CR's, (b) impossibility of predicting the course of CR generalization to the light patterns from either the "principles of patternization" of Wertheimer or the simple CR generalization of Pavlov, (c) the determining roles of ratios in the CR transfer from musical intervals.

Phonetographic Generalization (47, 49)

Four subjects were used. The conditioned stimuli were single English words, and the generalization stimuli were words related to the conditioned words in sound and spelling, a relatedness named by the writer phonetographic. A crude CR gradient was found here. Thus, the mean generalization to homophones (*urn*—"earn"; *style*—"stile"; *surf*—"serf"; *freeze*—"frieze") was 37 per cent, while the generalization from *flower* to "glower" was 35.1 per cent, from *dark* to "mark," 31.6 per cent, from *mock* to "dock," 27.2 per cent, from *flower* to "shower," 20.2 per cent, and from *day* to "may," 19.6 per cent. The *t*'s here are significant with *P* equalling .05, if the differences are more than four per cent.

Semantic Generalization (49)

Four to 11 subjects were used, and the conditioned stimuli were single English words. The generalization stimuli were word related to the conditioned words semantically; synonyms, contrasts, coordinates, supraordinates, subordinates, whole-part's, part-whole's, and so forth. The mean generalization to the synonyms was 59 per cent, while the generalization to the other word categories was *the greater the higher the frequency of such word categories in free association tests and the faster their reaction time in controlled association tests* (61, 32). However, this relationship held only if the relatedness between the conditioned and

the generalization words was classified according to "relatedness of conditioned words to generalization words," and not as a "relatedness of generalization words to conditioned words"; e.g., only if a conditioned word of "dog" and a generalization word of "animal" or a conditioned word of "flower" and a generalization word of "petal" were classed respectively as subordinates and as part-whole's, and not as supraordinates and as whole-part's. The significance of this finding will be discussed in a later section.

Effects of Mental Sets upon Semantic CR Generalization (50)

This study was largely an extension of the one preceding. Before being tested for generalization to various word categories, nine subjects were divided into three equal and equated groups. Group C_1 practiced *corresponding* word categories in controlled association tests, e.g., practiced supraordinates before being tested for CR generalization to supraordinates. Group C_2 practiced *converse* word categories, e.g., practiced subordinates before CR tests for supraordinates, while Group C_3 was not subjected to any such preliminary practice. As a result, Group C_3 , the control group, showed a mean CR generalization of 34 per cent. Group C_1 that practiced corresponding word categories manifested a mean generalization of 54 per cent, and Group C_2 that practiced converse categories, a mean of 25 per cent. The t 's here are significant for all three group differences with P equalling .01.

Effects of Cognitive (Knowledge of Stimulus Relations), Voluntary-Facilitatory, and Voluntary-Inhibitory Attitudes

Twelve subjects, divided into four equal and equated groups, were, first, conditioned to the tone C on a Stoelting *just* harmonium and to the word *flower*; and, then, tested for generalization to (a) tones $F, B, e, a, d^1, g^1, c^{\sharp 2}, f^{\sharp 2}$, and to (b) words "flour," "glower," "shower," and "scour." The cognitive attitudes for tone generalization were induced by the instructions of: "When you hear a tone in this session, try to think of its relatedness to the tone with which we experimented in the last two sessions. Try to think whether the tone that you will hear is higher or lower and how much higher and how much lower than the tone with which we experimented." The voluntary-facilitatory attitudes were instilled by: "I'd like to see how well you can control reactions. You were conditioned to secrete saliva to a low tone on this instrument [conditioning explained at some length]. From now on, you will hear eight higher tones, each about two or three musical intervals higher than the other, and it is your task to secrete most saliva to the tone closest to the conditioned tone—less, though, than to the conditioned tone itself—and least saliva to the tone most removed from the conditioned tone [all eight tones are now demonstrated]. In other words, I would like you to build up a saliva scale that will correspond as closely as

possible to the scale of tones here." The instructions for the voluntary-inhibitory attitudes differed from those for the voluntary-facilitatory attitudes in that the subjects were told to "secrete an equal amount of saliva to each tone irrespective of its pitch or quality." The instructions for word generalizations were similar to those of tone generalization except that the subjects were also told "not to pay attention to word relatedness in meaning but only to that of sound and spelling." Finally, there was a control group whose subjects were not given any instructions about the generalization.

The results are presented in Table VI. Space forbids a detailed discussion of the table. But a summary statement can be made: *within the limits of the present experiment cognitive and voluntary attitudes modified only a little the subjects' gradients of CR generalization.*

TABLE VI

EFFECTS OF COGNITIVE (KNOWLEDGE OF STIMULUS RELATIONS), VOLUNTARY-FACILITATORY, AND VOLUNTARY-INHIBITORY ATTITUDES UPON STIMULUS GENERALIZATION OF SALIVARY CONDITIONING IN HUMAN SUBJECTS

The subjects were 12 college undergraduates and the conditioned stimuli were (a) the tone C on a harmonium and (b) the word *flower*. Each entry is a mean of six determinations, two from each of the three subjects. VF = voluntary-facilitatory attitude, VI = voluntary-inhibitory, KR = knowledge of stimulus relations, UI = uninstructed.

Induced Attitude	Per Cent of Generalization Generalization Stimuli Tones								Reliable Gradient Steps (estimated)
	F	B	e	a	d ¹	g ¹	c [#]	f [#]	
UI	67	49	54	42	40	32	41	38	Three
VF	64	52	44	50	39	40	30	24	Three or Four
VI	78	72	76	74	56	58	42	48	Two or Three
KR	62	54	50	41	34	40	32	31	Three or Four
Words									
	Flour	Glower	Shower	Scour					
UI	39	29	32	19					
VF	62	68	52	23					
VI	68	72	60	48					
KR	54	41	45	26					

The Experiment with Transliterated Russian Words

Nine subjects, unfamiliar with Russian, were divided into three equal and equated groups and conditioned to four transliterated Russian words: BUKVA, DOLGO, KUPIT, and SMESHNOY. Group N was not told the meanings of the words. Group G was told the meanings of

the words before being tested for generalization, while Group CG was given one set of meanings before the primary training of the CR, and another reverse set of meanings before the generalization tests (having been assured that the second set was the correct one). The subjects in Group N did not, of course, show any significant semantic CR generalization. But the important point brought out by the experiment was that the course and range of the semantic generalization was determined entirely by the meanings given to the Russian words before the generalization tests. *The meanings imparted prior to the original training of the CR had no effect upon the subsequent semantic generalization of the CR to the verbal stimuli.*

Summary of Razran's Results

Besides corroborating in the main the findings of the Yale and the Pavlov laboratories (as analyzed by the writer), the writer's results indicate that:

1. With human subjects, gradients of CR generalization, though very crude ones, are more evident with verbal than with sensory conditioned stimuli.
2. With musical intervals as the conditioned stimuli, common ratios determine much more the course of CR generalization than common constituent tones, either fundamentals or partials; with flashes of lights as the conditioned stimuli, the spatial patterns of the lights are more determining than the colors of the lights in the patterns; and with verbal conditioned stimuli, relatedness in meanings is more significant than relatedness in sound and spelling.
3. Mental sets are effective in determining the general direction and magnitude of CR generalization but do not seem to be of much significance—without special training—in creating or in counteracting the *gradients* of the generalization.
4. The course and very existence of CR generalization is more likely a function of the subsequent testing for generalization than of the original training of the conditioning.⁸

THE TOTAL EVIDENCE AND THE LASHLEY-WADE INTERPRETATIONS

Total evidence adduced heretofore is, in the writer's estimate, discordant with either (a) the views of Pavlov that CR generalization is a function of the undulation of cortical excitation or (b) the views of Hull that the generalization gradient is a monotonic decreasing function of the magnitude of the differences between the conditioned and the generalization stimuli. But the possible concordance of the evidence with

⁸ In one study by the writer (50), the CR generalizations of four subjects, who were conditioned to controversial sentences such as "socialism is desirable," seemed to be affected by the socio-political views of the subjects, while the conditioning itself showed no such differences. However, these results were obtained from a very small number of subjects and must be considered very tentative.

the Lashley and Wade interpretations—outlined in the Introduction—deserves careful examination.

Generalization as a failure of association. There is no doubt in the writer's mind that a good deal of presumed evidence of CR generalization is due to a failure of association. To Lashley and Wade's insightful observations, he would add the following experimental support:

1. In six different experiments from Pavlov's laboratory (24, 27, 37, 53, 62, 63) it was shown that dogs conditioned to a compound stimulus—light and taction, light and sound, two different sounds or lights or tactile stimuli—had usually failed to produce a CR, when only one of the stimuli, supposedly the weaker one, was presented alone. There was no generalization to these components of the compound stimulus, even though they were physically present during the original CR training and even though they obviously were physically similar to the compound stimulus (partial identity.) And the only sensible interpretation of this phenomenon would seem to be that the animals failed to associate these particular component stimuli with the total CR situation, that the stimuli were not "attended to," were present but ineffective. Pavlovian explanations that in such cases the weaker stimuli become inhibited by the stronger ones contradict their own doctrine that all aspects of a stimulus are associated with the conditioned response. Or, to put it differently, if weaker stimuli are inhibited by stronger stimuli during primary conditioning, why do not the weaker generalization bonds become inhibited by the stronger conditioned bonds, and why is there generalization altogether?

2. In a number of experiments from Pavlov's laboratory, partially decorticated or badly injured or specially drugged dogs are often reported to manifest a great deal of CR generalization. And a widening of generalization is also often described in old dogs, young puppies, and in lower animals. Inasmuch as in all these case associative capacity is no doubt reduced, an interpretation of their generalizations in terms of failure of association seems very reasonable.

However, against these two considerations the writer would pit four others: viz.:

1. Partially decorticated and injured, and old dogs have also been reported by Pavlov and his students as having reduced greatly the scope of their generalization.

2. The frequent increase of CR generalization in the initial stages of training the CR—which the writer believes to be an established fact—does not go well with a "failure of association" interpretation.

3. Failure of association could hardly apply to the generalizations obtained by Grandine and Harlow (9) and by Grice (10) who used a Lashley-Wade technique.

4. Failure of association could not explain the generalizations in the writer's experiments in which the subjects attended very minutely to the stimuli, and probably not too well the generalizations in other experiments with human subjects.

In other words, while failure of association may account, in the

writer's estimate, for a good deal of *reported* CR generalization, it does not explain *all* CR generalization reported. Probably Lashley and Wade did not themselves mean it to do so.

Generalization as a product of variable stimulus thresholds. Again, it seems well probable that some data on CR generalization are to be attributed to variable stimulus thresholds. In spatial generalization of taction, Loucks (30) actually demonstrated what he called peripheral mechanical irradiation by stimulating a point not too far from the conditioned spot with a "pricker" obtained from Pavlov's laboratory. And the writer pointed out earlier in this article the probable masking effects of intensity in CR generalization. Still, there is a limit to variable thresholds, and the writer doubts very much whether such a concept could be of any value in interpreting CR generalization to stimuli far removed from the conditioned stimulus, inter-dimensional and inter-sensory generalization, generalization by the Lashley-Wade technique, and generalization in the writer's experiments.

Generalization along discriminated dimensions. The main argument here is the Lashley and Wade statement, quoted earlier, that "It [true generalization] does not occur in conditioning to a single stimulus but is somehow a function of differential training with two or more stimuli on the same dimension" (p. 82). This apparently means that, according to Lashley and Wade, whatever CR generalization cannot be explained by "failure of association" or by "variable stimulus thresholds," its cause must be sought in the reactional biographies of the stimuli involved, a laudable but difficult matter. To be sure, reactional biographies are very important, and the writer could point to the experiments by Prokofiev and Zeliony (38) and particularly to the one by Brogden (6) on sensory pre-conditioning which show the interactions of sensory stimuli in animals without even reinforcement; as well as to the effects of previous experience on CR gradients in his own experiments. Yet in the 67 studies from Pavlov's laboratory which the writer analyzed he ruled out results with dogs that had had, as far as could be ascertained, some previous experience with the generalization stimuli. And in the writer's own experiments, discriminative sets modified but did not wholly change the character of the subjects' CR gradients. Furthermore, on general grounds and analogies, Lashley and Wade's total denial of the existence of a true single-stimulus absolute value generalization is not wholly borne out by their own statement. They state: "... memory for relations is much more permanent than memory for absolute properties" (p. 86), which to the writer means that transposition is more significant than generalization, a view with which the writer tends, in a large way, to agree (cf. 44). But to the writer the statement also means an admission of the existence of memory for absolute prop-

erties and, by analogy, the existence of a true single-stimulus CR generalization.⁹

THE WRITER'S OWN INTERPRETATIONS

The writer's own interpretations of CR generalization rest upon three considerations: a *bifactoral theory*; a "subsequent testing" hypothesis; a doctrine of a qualitative, categorizing, "rating scale" type CR gradient.

The bifactoral theory. The bifactoral theory of CR generalization maintains that there are two kinds of generalizations: (a) *pseudo-generalization* and (b) *true generalization*. Pseudo-generalization, well accounted for by Lashley and Wade's interpretations, correlates, for the most part, negatively with organismic capacity. It predominates in the CR generalizations of lower animals (fish, guinea-pigs), in partially decorticated or very young or very old higher animals, in states of fatigue and inattention, and in some stages of CR training (initial, over-training). True generalization, on the other hand, is a positive capacity of the organism, an ability to generalize *absolute* characteristics of stimuli or objects. It seems to increase with maturity, alertness, post-operative recovery from cortical extirpations, administration of some drugs, and other capacity-enhancing influences.¹⁰ Analogically, perhaps, the relation of pseudo-generalization to true generalization is not unlike that of the undifferentiated total action of the young foetus to the structured whole activities of the fully developed individual.

The subsequent testing hypothesis. The "subsequent testing" hypothesis asserts that CR generalization develops, not during the original training of the conditioned stimuli, but during the subsequent testing of the generalization stimuli. The writer agrees here fully with Lashley and Wade that "there is no 'irradiation' or spread of effects during primary conditioning" (p. 74) and disagrees with the assumptions of Pavlov and Hull and Spence that during original CR training some sort of generalization bonds, latent but later ready to function, are automatically developed. He believes that such assumptions are not only cumbersome physiologically and superfluous logically, but also

⁹ The writer agrees also with Lashley and Wade that the "basic nervous mechanism is one of reaction to ratios of excitation" (p. 86). But he believes that some ratio of excitation is also involved in simple conditioning, a ratio between the reaction-patterns of the conditioned and the unconditioned stimuli.

¹⁰ The writer regrets not to be able to discuss here the evidence for these assertions. One reason is lack of space. Another is that the evidence has not been fully analyzed yet, and that these assertions are thus based only upon impressions of reading Russian studies, and in the case of lower animals upon an earlier, not completely up to date, review by the writer (41) of the topic.

that they are contradicted by a series of his own experiments. According to his view, *all effects of generalization are generated during tests of generalization*.¹¹

The doctrine of a qualitative, categorizing, "rating scale" type of CR gradient. This doctrine is in a way the crux of the writer's interpretations. It is based upon a statistical and a logical analysis of total evidence which, in the writer's estimate, demonstrates that there is a true CR gradient, but that this gradient is very qualitative and very crude, consisting of only a few steps, perhaps more steps in human beings than in dogs, but few just the same. Apparently, when human beings or dogs that have been conditioned to some stimulus or object are confronted with some new non-conditioned but in some way related stimulus or object, they categorize or rate the new stimulus on some sort of crude similarity-dissimilarity scale. With human subjects, introspections actually reveal such categorizing attitudes as "similar," "very similar," "not so similar," "somewhat similar," "dissimilar," "very dissimilar," and the like—attitudes that apparently control or even initiate the generalization responses. And it is the writer's contention that some such categorizing behavior is operative also in animals.

This categorizing is thought by the writer to be very dynamic and changeable and *varying much more with the organic dimensions of the organism than with the physical dimensions of the external stimuli in the CR situation* (data deny significant variations with the latter). Its neurophysiology is regrettably obscure, as regrettably as the neurophysiology of learning in general.¹² But it should not be difficult to conceive the organic dimensions with which it might correlate: perhaps the movement-produced stimuli of Guthrie, or the situation-sets of Woodworth, or even the means-ends-capacities of Tolman. At any rate, qualitative step-like CR gradients and only such gradients are objectively demonstrable. So that those of the writer's colleagues who fear anthropomorphism more than zoomorphism need accept only the type of gradient and not its human analogy. While to others, *rating scales in dogs* should be no less welcome than *hypotheses in rats*.

GENERAL SUMMARY

Four views of CR generalization have been considered:

1. *Cortico-physiological* (Pavlov, Bekhterev): CR generalization is a function of a wave-like irradiation of cortical excitation. It posits a quantitative continuous CR gradient.
2. *Physico-behavioral* (Hull, Spence): CR generalization is a monotonic de-

¹¹ It might be worth considering a similar hypothesis for transposition and for more complex transfer phenomena.

¹² The writer is much more in accord with Loucks than with Skinner and Hull about the need and value of correlating behavior with neurophysiology; but he does defer to Lashley about the paucity of our knowledge in this area.

creasing function of the magnitude of the differences between the conditioned stimulus and the generalization stimuli. It, too, posits a quantitative continuous CR gradient.

3. *Failure of association-transposition* (Lashley and Wade): some reported data on CR generalization are due to failure of association; others represent transpositions of discriminated dimensions. It denies the existence of true CR generalization.

4. *Categorizing-rating* (Razran): While recognizing the roles of failure of association and transposition, it affirms the existence of a true CR generalization which is attributed to the organism's categorizing or rating of related stimuli on some sort of crude similarity-dissimilarity scale. It posits a very crude and qualitative CR gradient.

Adduced evidence favors the fourth view and is clearly in disagreement with the first two views.¹³

BIBLIOGRAPHY

1. ANREP, G. V. The irradiation of conditioned reflexes. *Proc. Roy. Soc. London*, 1923, **94B**, 404-425.
2. BASS, M. J., & HULL, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47-65.
3. BEKHTEREV, V. M. *Psychology, reflexology, and Marxism*. Leningrad: GIZ, 1925.
4. BEKHTEREV, V. M. *General principles of human reflexology*. London: Jarrolds, 1933.
5. BERITOFF, J. S. *Individually-acquired activity of the central nervous system*. Tbilisi: GIZ, 1932.
6. BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323-332.
7. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209-226.
8. CARR, H. A. The laws of association. *Psychol. Rev.*, 1931, **38**, 212-228.
9. GRANDINE, L., & HARLOW, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. Psychol.*, 1948, **41**, 327-338.
10. GRICE, C. R. The effect of differential training to single stimuli upon the acquisition of a size discrimination habit. *Amer. Psychologist*, 1948, **3**, 242.
11. GUTHRIE, E. R. Conditioning as a principle of learning. *Psychol. Rev.*, 1930, **37**, 412-428.
12. GUTHRIE, E. R. Pavlov's theory of conditioning. *Psychol. Rev.*, 1934, **41**, 199-206.
13. GUTHRIE, E. R. *The psychology of learning*. New York: Harpers, 1935.
14. HILGARD, E. R. *Theories of learning*. New York: Appleton-Century, 1948.
15. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
16. HOVLAND, C. I. The generalization of

¹³ The writer regrets the omission, through an oversight, of the pertinent study by Grant and Dittmer (*J. exp. Psychol.*, 1940, **26**, 299-310). However, only 2 of the 31 subjects in the study showed generalization gradients without reversals while 17 subjects had mean CRs to generalization stimuli greater than to the conditioned stimulus, so that the results of these experimenters are in line with the present analysis. Grant himself is of course by no means an adherent of Pavlovian irradiation and—the writer ventures—of Hullian generalization.

- conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, 17, 125-148.
17. HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, 51, 279-291.
 18. HOVLAND, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, 21, 261-276.
 19. HULL, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, 46, 9-30.
 20. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
 21. HULL, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, 54, 120-134.
 22. HULL, C. L., et al. *Mathematico-deductive theory of rote learning*. New Haven: Yale Univ. Press, 1940.
 23. HUMPHREYS, L. G. Generalization as a function of the method of reinforcement. *J. exp. Psychol.*, 1939, 25, 141-158.
 24. KASHERININOVA, N. A. *Contributions to the study of salivary conditioned reflexes in response to tactile stimuli in the dog*. Thesis, St. Petersburg, 1908.
 25. KONORSKI, J. *Conditioned reflexes and neuron organization*. Cambridge: Cambridge Univ. Press, 1948.
 26. KRASNOGORSKI, N. I. Bedingte und unbedingte Reflexe in Kindesalter und ihre Bedeutung für die Klinik. *Ergebn. inner. Medis. Kinderhk.*, 1931, 39, 613-730.
 27. KUPALOV, P. S., & GANTT, W. H. The relation between intensities of conditioned stimuli and magnitudes of conditioned reflexes. *Trud. Fiziol. Lab. Pavlova*, 1928, 2, 1-12.
 28. LASHLEY, K. S. Learning: I. Nervous mechanisms in learning. In C. Murchison (Ed.), *The foundations of experimental psychology*. Worcester, Mass.: Clark Univ. Press, 1929. Pp. 524-563.
 29. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, 53, 72-87.
 30. LOUCKS, R. B. An appraisal of Pavlov's systematization of behavior from the experimental standpoint. *J. comp. Psychol.*, 1933, 15, 1-45.
 31. LOUCKS, R. B. Reflexology and the psychobiological approach. *Psychol. Rev.*, 1937, 44, 320-338.
 32. MAY, M. A. The mechanism of controlled association. *Arch Psychol.*, N. Y., 1917, No. 39.
 33. PAVLOV, I. P. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press, 1927.
 34. PAVLOV, I. P. *Lectures on conditioned reflexes: twenty-five years of objective study of the higher nervous activity (behavior) of animals*. New York: International Publishers, 1928.
 35. PAVLOV, I. P. The reply of a physiologist to psychologists. *Psychol. Rev.*, 1932, 39, 91-127.
 36. PAVLOV, I. P. *Conditioned reflexes and psychiatry*. New York: International Publishers, 1941.
 37. PERELZWEIG, I. *Materials to the study of conditioned reflexes*. Thesis, St. Petersburg, 1907.
 38. PROKOFIEV, G., & ZELIONY, G. Des modes d'associations cérébrales chez l'homme et chez les animaux. *J. de Psychol.*, 1926, 23, 1020-1028.
 39. RAZRAN, G. Theory of conditioning and related phenomena. *Psychol. Rev.*, 1930, 37, 25-43.
 40. RAZRAN, G. Conditioned responses in

- children: a behavioral and quantitative critical review of experimental studies. *Arch. Psychol. N.Y.*, 1933, 23, No. 148.
41. RAZRAN, G. Conditioned responses in animals other than dogs: a behavioral and quantitative critical review of experimental studies. *Psychol. Bull.*, 1933, 30, 261-324.
 42. RAZRAN, G. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol., N.Y.*, 1935, 28, No. 191.
 43. RAZRAN, G. Transposition of relational responses and generalization of conditioned responses. *Psychol. Rev.*, 1938, 45, 532-538.
 44. RAZRAN, G. Studies in configural conditioning: VII. Ratios and elements in salivary conditioning to various musical intervals. *Psychol. Rec.*, 1938, 2, 370-376.
 45. RAZRAN, G. The nature of the extinction process. *Psychol. Rev.*, 1939, 46, 264-297.
 46. RAZRAN, G. A simple technique for controlling subjective attitudes in salivary conditioning of adult human subjects. *Science*, 1939, 89, 160-161.
 47. RAZRAN, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, 90, 89-90.
 48. RAZRAN, G. Studies in configural conditioning: V. Generalization and transposition. *J. genet. Psychol.*, 1940, 56, 3-11.
 49. RAZRAN, G. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *J. exp. Psychol.* (Oct. 1949, in press).
 50. RAZRAN, G. Sentential and propositional generalizations of salivary conditioning to verbal stimuli. *Science*, 1949, 109, 447-448.
 51. RAZRAN, G. Some psychological factors in the generalization of salivary conditioning to verbal stimuli. *Amer. J. Psychol.*, 1949, 62, 247-256.
 52. RAZRAN, G., & WARDEN, C. J. The sensory capacity of the dog as studied by the conditioned reflex method (Russian schools). *Psychol. Bull.*, 1929, 26, 202-222.
 53. RICKMAN, V. The problem of the intensity of conditioned reflexes. *Trud. Fiziol. Lab. Pavlova*, 1928, 2, 13-21.
 54. ROBINSON, E. S. *Association theory today*. New York: Appleton-Century, 1932.
 55. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, 43, 427-449.
 56. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.
 57. SPENCE, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, 47, 271-288.
 58. WARREN, H. C. *A history of the association psychology*. New York: Scribners, 1921.
 59. WATSON, J. B., & RAYNOR, R. Conditioned emotional reactions. *J. exp. Psychol.*, 1920, 3, 1-14.
 60. WICKENS, D. D. Studies of response generalization in conditioning: I. Stimulus generalization during response generalization. *J. exp. Psychol.*, 1943, 33, 221-227.
 61. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
 62. ZELIONV, G. *Reactions of dogs to auditory stimuli*. Thesis, St. Petersburg, 1907.
 63. ZELIONV, G. Contribution a l'analyse des excitants complexes des reflexes conditionnels. *Ark. biol. Nauk*, 1910, 15, 437-453.
 64. Trudy Fiziologicheskikh Laboratorii Akademika I.P. Pavlova, 1924-1941, 1-10.

Received March 6, 1949.

PSYCHOTHERAPY AS A PROBLEM IN LEARNING THEORY¹

EDWARD JOSEPH SHOBN, JR.

State University of Iowa

It has become increasingly apparent that clinical psychologists are more and more drawing psychotherapy into their compass of activities. If this enlargement of scope is to be something more than a trading of one's psychological birthright for a share of psychiatric pottage, it would seem imperative that the therapeutic functions of the psychologist be regarded from the point of view of research as well as from that of practice. As Sanford (47) puts it,

What should be of great help to us here is our training in scientific method and our tradition of research-mindedness. It would be hard to name an area in which research is more needed than it is in therapy, or an area in which what is being done lags further behind what might be done. . . . And one might say, furthermore, that it is primarily up to the psychologist to perform this needed research.

The difficulties in the way of such inquiry, however, are enormous, as is well attested to by the paucity of investigations of the therapeutic process in terms of the problems, techniques, and concepts common to general psychology. The nature of some of these barriers to psychological research on a matter of such importance probably merits some brief attention.

In the first place, there are situational deterrents to research in psychotherapy. Counseling² usually takes place in a "service" setting and is consequently seldom subject to the kinds of exact manipulation required by rigorous experimentation. Often, attempts to control various factors in the therapeutic set-up give rise to serious ethical problems concerning the relationship of the therapist and his agency to their clients, and certainly the pressure of the demand for counseling services frequently conflicts with the requirements of a research program. Sec-

¹ This article represents a revision and extension of an earlier attempt (51) to conceptualize psychotherapy in terms of systematic behavior theory. Acknowledgment must be made to a number of people, foremost among whom is Dr. O. H. Mowrer, who, though he may recognize some of his ideas in the ensuing pages, must not be held responsible either for their form or for the uses to which they are put. Others are Dr. Kenneth Spence and Dr. I. E. Farber, of the University of Iowa, who have been invaluable sources of stimulation and instruction but who are absolved from any responsibility for what is here said.

² The terms counseling and psychotherapy are here used interchangeably without regard for any of the distinctions they are sometimes employed to convey.

ondly, the problem of complexity gives one pause. Psychotherapy is a form of social interaction, an active social situation, in which many subtle, difficult-to-isolate aspects of the personalities of *both* patient and counselor must be taken into consideration. The therapist is not merely the wielder of some supposedly meliorative technique but is deeply involved as a personality in the counseling process. Thus, the psychology of the psychologist, as well as the psychology of the patient and the nature of the therapeutic method, enters into the determination of the therapeutic end product. Third, there are personnel problems militating against effective research in psychotherapy. Psychologists most familiar with the therapeutic process are seldom well schooled in the experimental and conceptual skills basic to fruitful investigations in general psychology, whereas those who are best equipped technically and conceptually as research workers are generally rather untutored in therapeutic techniques, are unfamiliar with clinical material, and are frequently repelled by the admittedly gross and somewhat nebulous notions clinicians use in their efforts to conceptualize the complex phenomena with which they work. In sum, the situational lack of amenability of psychotherapy to experimental inquiry, the enormous complexity of the factors entering into the counseling process, and the differences in training and interest between clinical and laboratory workers all tend to impede a *rapprochement* between psychotherapy and the research functions characteristic of general psychology.

In spite of these difficulties, there is one slender lead that might be profitably followed in the attempt to provide a basis for the conceptualization and investigation of psychotherapy as a problem in general psychology. This is the widespread recognition that psychotherapy is essentially a learning process and should be subject to study as such.

This point of view is not only in harmony with the general conception of counseling as a conversation or series of conversations between two persons, therapist and patient, the goal of which is to resolve the conflicts, reduce the anxiety, or somehow modify the behavior of the latter—a conception which clearly implies learning; it has been more or less clearly so verbalized by a number of clinical workers. Cameron (4) sees the desideratum of counseling as the patient's "acquisition of normal biosocial behavior," a statement which definitely implies the learning of new ways of reacting as a function of the therapeutic process. Alexander and French (1) advance as a basic therapeutic principle the reexposure of the client, within the favorable circumstances of psychotherapy, to emotional situations with which he was unable to deal in the past. Presumably, the justification for such a reexposure rests on

the hypothesis that its occurrence "under more favorable conditions" in some way permits the patient to learn more adequate ways of coping with such experiences. Rogers (45) describes the therapeutic process as a freeing of the "growth capacities" of the individual which permits him to acquire "more mature" ways of reacting. If "growth" in this context means (as it must) something more than physiological maturation, and if it is not to be lumped with the old and rather mystic homeopathic notion of the *vis medicatrix naturae*, it must refer to the client's acquisition of new modes of response. Such new modes of response are "more mature" because for a given patient they are less fraught with anxiety or conflict. Thus, Rogers is actually talking about psychotherapy as a learning process. White (64) insists that, "Psychotherapy is designed to bring about learning . . ."; and Darley (7) argues that unless the process of learning in counseling is demonstrated, it is not legitimate to infer that the modifications of behavior that may occur during or following therapy are necessarily outcomes of therapy.

In spite of this widespread acknowledgement of psychotherapy as a learning process, there have been few attempts (11, 25, 49, 50) systematically to formulate therapy in terms of learning theory. This paper represents a tentative, apologetically offered effort to construct a learning-theory interpretation of counseling that will help to narrow the gap between practitioner and researcher, clinician and experimentalist, and to encourage some much needed investigation.

COMMON FACTORS IN SCHOOLS OF PSYCHOTHERAPY

When one surveys the various theories and practices of psychotherapy in an effort to find those common factors which a learning-theory interpretation of the counseling process must cover, it appears possible to make four summarizing general statements:

1. All schools of psychotherapy can with some justice claim cures (46). Notable successes seem to be the common property of virtually all forms of counseling from moral suasion through non-directive therapy to psychoanalysis.

2. Clinical patients,³ in spite of their enormous differences, tend to present a similar problem in that one of their primary motivations is anxiety and much of their non-integrative or "symptomatic" behavior is maintained on the basis of anxiety reduction.

3. The goal common to most psychotherapies is the modification of the client's underlying anxiety. This is related to the hypothesis that once his motivation is altered, the overt habit structures of the patient will change.

³ The "clinical patients" spoken of in this paper include only those classifiable as neurotic or "maladjusted." Nothing said here is meant to apply to psychotics, psychopaths, or behavior problems associated with endocrine disturbances or lesions of the central or autonomic nervous systems.

4. Finally, all types of counseling employ the techniques of the *therapeutic relationship*, the unique social situation that is formed when therapist and patient meet to discuss the problems of the latter, and of *conversational content*, that is, of talking about certain things within the therapeutic setting rather than others.

A word must be said about each of these four factors which seem to be common to the various forms of counseling, regardless of the doctrinal banners flown.

All schools report cures. If it is true that the proponents of various theories of psychotherapy all seem able to claim successes, and if it is true—as has often been pointed out—that successes are no proof of therapeutic theory, then it would seem to follow that an understanding of the counseling process would be furthered by giving more attention to the conditions under which the patient's learning of new modes of reaction takes place within the general clinical setting. If this is a fair notion, based as it is on the conception of therapy as a learning situation, it might be instructive to explore the points in common among the different approaches to counseling in terms of (a) the similarity of patients' problems, (b) the agreement among clinicians as to goals, and (c) the techniques common to nearly all therapeutic enterprises. Such an exploration might lead to a formulation of the learning process in counseling in terms of these three sets of information.

Similarities in clinical cases. While from the practical standpoint of dealing therapeutically with patients it is necessary to consider each case in all its uniqueness, from a theoretical point of view it is instructive to look for similarities. This amounts to asking the rather ambitious questions of (a) What constitutes the core of "neurosis" or "maladjustment"? and (b) What are the common problems faced by therapists in their contacts with patients? While no definitive answer can be given here, it is important to consider these issues as bearing on the goals and techniques employed by counselors of different theoretical persuasions and as factors to be accounted for in attempting to formulate a learning-theory interpretation of the therapeutic process.

A point on which there seems to be widespread agreement is, in Horney's (17) phrase, that "one essential factor common to all neuroses . . . is anxieties and the defenses built up against them." The phenomena clinically identified as feelings of insecurity, feelings of inadequacy, and guilt feelings are all variants of anxiety in the sense that they involve debilitating expectations of future punishment. Likewise, it would seem that the "phenomenological self-concept" of Combs (6) and Rogers (45) refers to little more than a patient's level of anxiety, guilt, or inadequacy, together with his verbalizations, accurate or otherwise, of his defenses against them.

To conceptualize anxiety usefully, it is necessary to discriminate between anxiety and fear or, as Freud (12, 13, 14) did, between neurotic

anxiety and objective anxiety. Fear may be thought of as an affective reaction proportionate to some external danger. Anxiety, on the other hand, differs from fear in at least two ways. First, if one asks a "neurotic" patient what he is afraid of, he will admit to being afraid but will generally have no idea of what the source of the possible danger might be. Anxiety may be aptly termed either a fear of "nothing" or a fear of something which is objectively irrelevant. Second, while both fear and anxiety are anticipatory states involving some kind of premonition of danger, the signal to which anxiety is a reaction is usually internal, some impulse to act in a way that has been forbidden. An illustrative case may clarify this point.

E. B., a 24-year-old male undergraduate veteran, despite slightly better than average academic ability, is making poor grades and is in danger of being dismissed from his university. He complains of being "unable" to study, feelings of inferiority in social groups, and serious doubts as to both his intellectual and social adequacy. He has some guilt feelings about having transferred from a pre-medical curriculum to English, because his parents are quite eager for him to become a physician. His father is a farmer who has been quite successful financially and in community politics, and who has been highly ambitious for his son. He has imposed very high standards of attainment on the boy, has been quite strict and stern with him and has had a number of set ideas which he felt that the youngster should accept and act upon "for his own good." Any deviation on the part of the patient from the parentally prescribed ways of doing was met with severe punishment, the verbal part of which usually consisted in a variety of changes rung on the theme of the boy's worthlessness and a series of predictions that he would come to no good end. In short, any self-initiated activity—behavior which the parents themselves did not lay out—was fraught with danger. When the boy began counseling, he was squarely on the horns of a dilemma: unable to meet parental demands for a variety of reasons, he was also unable to initiate any divergent plans of his own without experiencing a flood of anxiety, i.e., anticipations of parental punishment.

This, if it is acceptable, leads to a general formulation of non-integrative or neurotic behavior. Anxiety has repeatedly been shown to have drive properties (29, 32), and on the basis of the anxiety drive, individuals who are maladjusted seem to develop various overt reaction patterns that become stable according to the degree to which they reduce the anxiety. This statement in terms of contemporary reinforcement theory (19) is quite in keeping with Freud's (12) idea of the interchangeability of anxiety and symptom, by which he means that through the formation of symptoms the patient protects himself from anxiety attacks. Anxiety is allayed by some anxiety-reducing symptom; if the symptomatic behavior is somehow eliminated, the anxiety returns. On the basis of this notion it is possible to define a neurosis or a maladjustment in terms of behavior which serves to reduce anxiety directly *without altering the conditions which produce the anxiety*. Freud consistently refers to anxiety as a signal of impending danger; the malad-

justed person is one who either consciously or unconsciously engages in acts which eliminate or neutralize the signal while leaving the objective danger unaffected. He is in the position of the motorist who shuts his eyes to warnings of dangerous curves, thus protecting himself from worry but leaving himself liable to serious accidents.

Such a conception permits an explanation of the curious observation that non-integrative behavior is at the same time self-defeating and self-perpetuating. It is self-defeating in that such behavior leads inevitably to further punishment: the motorist has accidents; the illustrative case suffers academic failures and social disarticulation through his avoidance of study to protect himself from the anxiety engendered by self-initiated activity and his withdrawal from social affairs to hide his "worthlessness." It is self-perpetuating because of the immediate reinforcement derived from anxiety reduction. Since the occurrence of a reinforcing state of affairs lies on the temporal gradient of reinforcement in greater proximity to the anxiety-reducing behavior than does the more remote punishment, the connection between the external and internal cues of anxiety and the non-integrative response tends to be strengthened (38).

A necessary concept in a theory of anxiety is that of repression. This notion refers to the exclusion from communicability (consciousness) of an impulse to act which has led to punishment. When a parent punishes a child severely for some tabooed act, the impulse to commit such an act becomes, through its association with the punishment, a stimulus for anxiety. One way by which the anxiety may be avoided is through repression—the exclusion from awareness of the impulse. If the repression is complete, there is a thorough-going allaying of anxiety, and the forbidden impulse no longer constitutes a problem.

Difficulty arises because repression is seldom if ever complete. The individual is constantly threatened by "a return of the repressed" (14) which touches off anxiety without the patient's being able to verbalize the cues for it. In short, the repressed impulse, although excluded from communicability, is still operative at subliminal levels. Why this should be true is something of a psychological mystery, although some light is shed upon it by investigations of punishment. Estes (9), for example, by a series of experiments has shown that punishment does not extinguish a response which has been positively reinforced. He concludes,

.... a response cannot be eliminated from an organism's repertoire more rapidly with the aid of punishment than without it. In fact, severe punishment may have precisely the opposite effect. . . . The punished response continues to exist in the organism's repertoire with most of its original latent strength. While it is suppressed, the response is not only protected from extinction, but it also may become a source of conflict. An emotional state, such as "anxiety" or "dread," which has become conditioned to the incipient movements of making the re-

sponse will be aroused by any stimuli which formerly acted as occasions for the occurrence of the response (pp. 37-38).

This provides a neat parallel to what is implied in the concept of repression.

In summary, then, one might say that clinical cases share in common (a) anxiety touched off by (b) un verbalized, unsuccessfully repressed impulses to act in ways that have met with punishment, and (c) persistent non-integrative behavior of many kinds, which reduces the anxiety but does nothing about eliminating its objective causes.

Common goals in psychotherapy. In spite of its non-integrative nature, overt neurotic behavior acquires remarkable persistence through anxiety-avoidance. This persistence is probably the factor most responsible for the failure and consequent elimination of clinical techniques aimed at the elimination of symptoms. Such a goal, in effect, defined psychotherapy as a process of robbing the patient of his defenses against anxiety without alleviating the unbearable state of dread. Since such an end is impossible of realization, advice, persuasion, exhortation, and suggestion have largely gone by the board in favor of methods which focus on the client's anxiety itself.

In other words, the goal of most modern psychotherapies is the modification of the emotional determinants of neurotic behavior. Thus, Alexander and French (1) speak of therapy as "a corrective emotional experience," which presumably results in a diminution of anxiety and a consequent elimination of persistent non-integrative behavior from the patient's repertoire. Likewise, White (64) points out that "Psychotherapy does not take place primarily in the sphere of intellect Its sphere of operation is the patient's feelings." The kind of learning with which counseling is concerned has to do chiefly with the alteration of motives and affective drives. This does not mean, of course, that the therapist is uninterested in his client's overt behavior; on the contrary, it is his job to help the patient alter it and achieve a repertoire of more integrative habits. But since this goal does not seem attainable through any kind of direct manipulation, the counselor generally works on the elimination of the basic anxieties, implicitly hypothesizing that once the drive conditions are changed, the neurotic behavior will show less strength.

Common tools in psychotherapy. From the standpoint of technique, there are two main aspects of the counseling process, common to all schools of psychotherapy. One is the unique *relationship* that develops between therapist and patient; the other is the *conversational content*, what they talk about during their sessions together. The proponents of different theories of counseling may emphasize one or the other of these factors, but both figure in their final formulations of therapeutic procedure. Thus, Williamson (66) and Kraines (24) stress the therapist's

obtaining personal information from the client so that the counselor may guide him somehow to a higher level of adjustment. In spite of this emphasis, both these clinicians devote a good deal of attention to the necessity of establishing and maintaining rapport or winning and retaining the patient's confidence. On the other hand, therapists like Taft (60), Allen (2), and Rogers (44) play up the quality of the counselor-client relationship and are concerned only secondarily with the conversational content aspect of therapeutic interviews. Nonetheless, they are quite insistent that the proper content of counseling contacts is the "feelings" of the patient rather than his overt behavior or his intellectualized beliefs.

What is this content factor in counseling? What are the areas of discussion between counselor and counselee? In line with the foregoing (although at variance with a widespread belief among laymen), therapeutic conversations are concerned with the patient's overt behavior only insofar as it bears on his covert reactions—the anxieties from which he suffers and against which he so non-integratively defends himself.

The client's anxiety (guilt feelings, feelings of inferiority, or inadequacy), then, constitutes the central topic of concern in psychotherapeutic interviews. But clinicians are also interested in the occurrences that engender anxiety. Especially are they interested in the formative past experiences⁴ which have been associated with anxiety, and they encourage patients to discuss such events and their reactions to them rather fully. Emphasis throughout seems to be more on the way the client feels about his experience rather than on the objective accuracy of his reportage.

Thus the conversational content of counseling consists chiefly in the discussion of the patient's anxieties and the conditions which either currently evoke them or seem to be causally linked in some historical sense to them.

The relationship aspect of therapeutic procedure has been recently most vigorously expounded by Rogers (44), Snyder (54), and other members (3, 6) of the so-called non-directive or client-centered school. Such a notion is, of course, by no means new to counseling technique. Freud (13) in stressing the idea of transference was talking about essentially the same thing: the basic role in psychotherapy of the affective bonds uniting client to counselor. In the case of orthodox

⁴ Even therapists like Rogers, who verbally disclaim any interest in personal history data, hardly prevent their patients' discussing past experiences. It would be revealing to go systematically through a series of electrically recorded non-directive interviews to see if the data collected fall very far short of affording a relatively complete case history. In a preliminary trial by the writer, using material collected from twelve sessions with one case, the greater part of a typical anamnestic form could be filled out from the transcriptions of the recordings.

psychoanalysis, transference refers to the displacement of childish attitudes from the analysand's past to the analyst, who becomes a substitute for the important previous objects of his patient's loves and hates. That such things do take place in psychotherapy is not questioned, but whether they *must* occur in just such a form for counseling to be successful may be doubted. For present purposes, it is merely necessary to establish the point that the relationship factor is inherent in the psychoanalytic approach to therapy. Cameron (4), writing from a point of view strongly influenced by Adolph Meyer, says,

... the acquisition of normal biosocial behavior may be greatly facilitated by the organization of a permissive situation, in which the patient has maximal opportunity to work through his attitudes and responses overtly in the presence of a skilled therapist. . . . The immediate goal of treatment in the behavior disorders is that of establishing a biosocial interrelationship . . . in which patient and therapist participate. The ultimate goal is that of making this interrelationship unnecessary and terminating it with benefit to the patient (pp. 576-577).

Dejerine and Gauckler (8) warn, "If . . . you have not been able to awaken a reciprocal sympathy in your patient, and if you have not succeeded in gaining his confidence, it is useless to go any further. The result that you will obtain will be worthless . . ." Sullivan (58) stresses the concept of parataxis and speaks of the psychiatrist's "participating helpfully in the life of the patient."

While there may be some important differences among the various points of view just touched on, it may be pointed out that there is virtually universal agreement among clinicians on the *importance* of the relationship; there is also high agreement on certain of its characteristics.

The most underscored aspect of the therapeutic relationship seems to be its warmth, permissiveness, and complete freedom from moralistic and judgmental attitudes on the part of the counselor. Far from being a coldly objective consideration of the patient's troubles, therapy necessarily involves a highly personal form of interaction in which the counselor is highly acceptant of the client's behavior, both overt and covert, within clearly defined limits.

Just what "acceptance" means has become somewhat clouded, and a word of clarification may throw some light on the dynamics of the counseling relationship generally. As Sullivan (58) points out, anything a patient feels, says, or does constitutes the data of the therapeutic enterprise. As is the case with data of any kind, one's first job is to understand; it is not to condemn, ignore, reject, or judge. Among such data are the feelings and attitudes that the counselee may develop toward the therapist and which, according to most clinical workers of whatever theoretical orientation, are intimately related to the success or failure of therapy. Here again an atmosphere free from censure or judgment but pervaded by sympathetic understanding is provided by

the counselor. On the other hand, acceptance does not imply approval of the client's feelings, attitudes, or overt behavior. This is not surprising since most clinical cases hardly approve of themselves, and their self-disapproval provides one of the most important aspects of the discomfort that brings them into therapy.

As can be inferred from the foregoing, the counseling relationship differs importantly from other forms of human interaction. In the first place, it is essentially one-sided in the sense that the therapist ordinarily says little about himself and that the changes effected within the context of the relationship are centered in the client rather than being a mutual modification. The exchange between counselor and patient, then, does not resemble that between friends in spite of the friendliness that generally permeates the relationship. Secondly, it is sharply limited in that the therapist's expressed interest in his client does not extend beyond the confines of the clinic. The two do not mingle socially, the clinician does not usually intercede for the patient in times of stress, and he generally does not become embroiled in attempts to manipulate the patient's environment. The therapist's office is designated as a place where one can come in perfect safety, free from threats and blame, to "think about" one's problems; but it is not a place where dispensations are sold or intercessions granted. Finally, there is a tacit agreement between therapist and patient that their connection is to be severed as soon as the patient feels free to go about his business without the counselor's support. In other words, the interest, acceptance, and "affection" of the therapist is there for the client to make capital of so long as he wishes it. Unlike non-clinical situations, there is no pressure on him to maintain the relationship out of politeness or any of the other social rules that more or less govern intimate relationships in society at large.

All this may be recapitulated by saying that the methods common to the various forms of psychotherapy involve (a) the formation of a special kind of personal relationship and (b) a conversation with the patient about his anxieties and the events which tend to produce them. As Finesinger (10) puts it, "Communication. . . and the physician-patient relation are the tools that must be adapted to the goals of psychotherapy."

The argument thus far, then, runs something like this: The common problem characterizing clinical patients is anxiety and the behavioral defenses built up against it. The goal of psychotherapy, regardless of the therapist's theoretical leanings, is to eliminate the anxiety and thereby to do away with the symptomatic persistent non-integrative behavior. To accomplish this goal, all therapists use the devices of conversing with the patient about his anxiety and the situations calling it forth both currently and historically, and forming a unique therapeutic

relationship. Since all psychotherapies seem to have successes to their credit and since psychotherapy seems to be a process whereby a patient learns to modify his emotional reactions and his overt behavior, it is hypothesized that therapy may be conceptualized from the point of view of general psychology as a problem in learning theory. Such a conceptualization must account for the changes that occur in counselees in terms of these factors that are apparently common to all forms of counseling. Before attempting such a conceptualization, it is necessary briefly to review the situation in learning theory.

MAJOR THEORIES OF LEARNING

One of the major issues with which learning theorists are concerned has to do with the conditions which are necessary if learning is to occur. Two points of view have gained the widest currency with respect to this question.

Reinforcement theory. The first is that of Clark Hull (19). Within Hull's system, learning is thought to proceed somewhat in this manner: When a motivated organism is subjected to stimulation—from either or both the stimuli associated with the motivating conditions themselves, as in hunger or pain, and those acting on it from the external environment—it tends to respond in a trial-and-error way. If, in the course of its trial-and-error behavior, the organism performs a response which is associated with the reduction of motivation, the probability of that response's occurring again under similar stimulus conditions is increased, or—to put it somewhat differently—the connection between the present stimuli and the response is strengthened. The central emphasis here is on the occurrence of drive reduction or a satisfying state of affairs, variously designated as the law of effect or the principle of reinforcement. As Miller and Dollard (31) succinctly sum it up: To learn, an organism must want something (be motivated in some way), notice something (be acted upon by stimulus cues from the external or internal environment), do something (perform a response or response sequence), and get something (experience a reduction in motivation).

Contiguity theory. Opposed to a reinforcement theory of learning is a point of view which holds that the basic condition necessary for learning is that of contiguity in experience. Tolman and Guthrie are perhaps the outstanding proponents of this theory, although they differ markedly in their conceptions of the nature of learning.

Tolman (61), taking his point of departure essentially from *Gestalt-theorie*, conceives of learning as the acquisition of information or cognitions about the environment. Various referred to as "sign-gestalt expectations," "sign-significate relations," and "hypotheses," these cognitions presumably have reference to knowledge which the organism acquires to the effect that a given stimulus or sign, if reacted to in a

given way by the organism, will lead to a spatially or temporally more remote stimulus or significate. The necessary condition for the acquisition of such "cognitive maps," as Tolman (62) has called them, is contiguity, the spatial and temporal patterning of stimulus events from sign to significate in the organism's experience. Aided by such secondary principles as recency, emphasis, and belongingness, the law of association by contiguity governs *learning*; learning—i.e., the acquired cognitive maps—together with the organism's needs and skills governs *performance*.

For Guthrie (15) learning is conceived as the acquisition of stimulus-response bonds as is the case with Hull. Unlike Hull, however, he holds that the occurrence of reinforcement is not a necessary condition for learning. Instead, he states that the principle governing learning is association by contiguity: "A stimulus pattern that is acting at the time of a response will, if it recurs, tend to produce that response." Simultaneity of stimulus cues and response is all that is required for the formation of new S-R bonds. Drive states or the existence of unconditioned stimuli are important only as "forcers" of the response to be learned, not as the basis of reinforcement in the Hullian sense.

The behavior with which the various proponents of these points of view have been concerned in their experimentation has consisted for the most part of skeletal muscle acts—maze running, problem-box solutions, conditioned leg flexions, etc. With this fact kept in mind, it seems fair to conclude that the reinforcement point of view seems to have something of an edge in predictive and explanatory utility over contiguity theory. O'Connor (40) has argued rather devastatingly against Guthrie's position by showing that it cannot accommodate the facts of delayed-reward learning. Likewise, Spence and Lippitt (56), Spence and Kendler (55), and Kendler and Mencher (22) have thrown serious doubt on the adequacy of Tolman's notion of contiguity in experience of sign, significate, and response as the essential and sufficient condition for learning.

Reinforcement theory, on the other hand, has demonstrated its utility in a variety of ways. Whiting (65) has conceptualized the socialization process in terms of Hull's notions. Miller and Dollard (31) have made some fruitful incidental remarks on cultural diffusion. Miller (30) has shown the adequacy of the scheme for explaining certain psychopathological phenomena. Loucks (26) and Loucks and Gantt (27) have applied evidence that strongly supports Hull's contention that the classical conditioning of skeletal muscle responses is merely a special case of learning according to the principle of reinforcement.

It is precisely at this point, however—in the conditioning of defense reactions—that the law of effect runs into difficulties. Hull (18) pointed

out this problem as early as 1929, referring to it as "the dilemma of the conditioned defense reaction." He then wrote,

For a defense reaction to be wholly successful, it should take place so early that the organism will completely escape injury, i.e., the impact of the nocuous (unconditioned) stimulus. But in case the unconditioned stimulus fails to impinge upon the organism, there will be no reinforcement of the conditioned tendency, which means one would expect that experimental extinction will set in at once. This will rapidly render the conditioned reflex impotent, which, in turn, will expose the organism to the original injury. This will initiate a second cycle substantially like the first which will be followed by another and another indefinitely, a series of successful escapes [from all contact with the noxious stimulus] always alternating with a series of injuries. From a biological point of view, the picture emerging from the above theoretical considerations is decidedly not an attractive one.

There is thus presented a kind of biological dilemma . . . (p. 511).

In other words, reinforcement theory finds it hard to explain how an organism can learn to avoid painful stimulation entirely, because if the painful stimuli do not act upon the organism's receptors, no drive is aroused to act as a basis for maintaining the defense reaction.

Mowrer and Lamoreaux (36), concerning themselves with this problem, resolved the dilemma by positing a conditioned fear reaction to the conditioned stimulus. According to their formulation, the conditioned stimulus has signal value, signifying to the organism an approaching danger and arousing in it those anticipations of punishment known as the secondary (acquired) drive of fear (anxiety). On the basis of this secondary drive, trial-and-error behavior occurs, out of which is differentiated, according to the principle of reinforcement, a response which reduces the fear and permits the organism to avoid or to minimize the painful unconditioned stimulus.

Such a resolution of the dilemma of the conditioned defense reaction, however, gives rise to another difficulty of comparable magnitude: How is the fear learned? If one holds to a thoroughly monistic reinforcement position, he is forced to say that the drive state of fear or anxiety is somehow "satisfying" or motivation reducing. Baldly, the reinforcement theorist is forced to hold that secondary drive arousal occurs on the basis of drive reduction. That this is certainly contrary to any kind of common sense consideration is immediately apparent, and it is difficult to see how an exchange of one drive for another—the situation which would obtain were the law of effect rigidly adhered to—could be of any biological benefit. This is particularly true when one recalls that many fears, especially neurotic anxiety, are much more debilitating than the objective conditions which generate them—witness the many

people who cannot bear to have dental work done or who refuse to see doctors.

Thus, a kind of *impasse* is reached. Reinforcement theory seems to account rather adequately for the acquisition of striped muscle acts; but at least in the conditioned defense situation—most germane to the clinical problems here under scrutiny—its adequacy is dependent on the operation of secondary motivational states, for the acquisition of which it is hard put to it to explain.

Two-factor theory. A number of writers have attempted to overcome this obstacle to efficient theorizing by formulating *two principles* to explain *two different kinds of learning*. Schlosberg (48) in 1937 expressed himself, on the basis of a long series of studies in his laboratory, as believing that there were two types of learning. One had to do with the acquisition of "diffuse, preparatory responses," by which he meant such things as changes in breathing, pulse rate, electrical skin resistance, body volume, voice pitch, and tonicity, which proceeds by "simple conditioning" or according to the principle of association by sheer contiguity. It will be recognized that these reactions are essentially those autonomically mediated viscerovascular reactions usually thought of as the basic physiological concomitants of emotion. The other type of learning which he felt it necessary to distinguish referred to the acquisition of more "precise, adaptive responses," withdrawal, flexion, or more generally defensive reactions, which are governed by the principle of "success" or reinforcement. These, of course, are the skeletal muscle acts which Hull's kind of theorizing seems to account for so admirably, whether the experimental situation be of the classical or instrumental kind of conditioning.

Skinner (53) in his 1938 volume made explicit a point of view at which he had hinted earlier (52). He distinguished between Type S conditioning as preparatory and Type R as consummatory, holding that the fundamental distinction rested on the event with which the unconditioned stimulus was correlated. In Type S the unconditioned stimulus is correlated with the conditioned stimulus, whereas in Type R it is correlated with the response. Skinner further says,

Most of the experiments upon skeletal behavior which have been offered as paralleling Pavlov's work are capable of interpretation as discriminated operants of Type R. . . . It is quite possible on the existing evidence that a strict topographical separation of types following the skeletal-autonomic distinction may be made (53, p. 112).

In this formulation, the same classification as that suggested by Schlosberg is implied. Autonomically mediated "emotional" reactions are learned on the basis of contiguity, whereas centrally mediated skeletal muscle responses are learned on the basis of reinforcement.

Razran (43) in 1939 offered a somewhat similar formulation, classifying learning according to what he called "quantitative" and "qualitative" conditioning, corresponding to learning without reinforcement and law-of-effect learning. He reports no evidence for the so-called qualitative conditioning of autonomic reactions, but does not say explicitly that quantitative conditioning applies exclusively to the acquisition of viscerovascular reactions. He does raise the issue of the differential importance of two events, the *application* or onset of the unconditioned stimulus and the *termination* of the unconditioned stimulus, for the conceptualization of types of learning.

More recently, Mowrer (34) has vigorously exploited the idea of a two-factor theory of learning to account not only for the learning of skeletal muscle responses but for the acquisition of secondary drives like fear and anxiety. He fully accepts the notion that striped muscle acts, mediated by the central nervous system, are learned, according to the principle of reinforcement, by virtue of their association with the *termination* of the noxious stimulation identified as motivational states. This is not only fully in keeping with Hull's position but is quite in line with Mowrer's own previous enthusiastic experimentation and theorizing as a monistic member of the reinforcement school (33). His new point of view, however, holds that smooth muscle and glandular "emotional" reactions, autonomically mediated, are acquired through their association with the *onset* of the paired unconditioned stimulus of pain and conditioned stimulus or signal. In other words, fear refers to the viscerovascular components of the pain response, conditioned to a substitute stimulus through the latter's contiguity with the *onset* of the action of a noxious adequate stimulus. He prefers to restrict the term *conditioning* to the learning of "emotional" reactions by contiguity and to use *problem-solving* to designate the learning of skeletal responses which "solve" the "problems" created by drives and which are acquired according to the reinforcement principle.

One of the points which must be made immediately with respect to two-factor theories such as these is aimed at the scotching of the criticism often (and fairly) leveled against attempts to account for learning in terms of multiple principles. Such attempts frequently permit the theorist to invoke whichever notion happens most easily to explain his data; he can explain everything but predict nothing. With the possible exception of Razran's, the two-factor formulations just reviewed are not liable to such an attack. While two principles are postulated, contiguity and reinforcement, two learning processes, one involving the viscerovascular system and the other the skeletal muscular system, are also suggested. The principle that governs one process may not be invoked to explain what occurs in the other. For either process, the theory is monistic and parsimonious and presumably subject to an *experimentum crucis*.⁵

Direct experimental tests of the two-factor theory are as yet few.

One study having an immediate bearing on the issue is that of Mowrer and Suter (37). These researchers argue that if the drive-termination theory of acquiring "conditioned" responses is valid, the response should become more readily connected with those stimuli present at the time of drive reduction. If, on the other hand, the drive-onset interpretation is correct, there should be no difference in the resulting learning curves. The rationale on which this deduction is based, of course, is that a conditioned stimulus (warning signal) *must* coincide with or approximate the turning on of the noxious unconditioned stimulus. If this contiguity with the *onset* of drive is all that is necessary for "conditioning" to occur, it should make no difference whether the conditioned stimulus overlaps with the *termination* of the unconditioned stimulus or not. Using an arbitrary running response as an index of fear and as their criterion of conditioning, Mowrer and Suter obtained experimental results confirmatory of their prediction: there was no difference in the curves of response acquisition between a group of rats trained under conditions where the conditioned stimulus overlapped and terminated with the turning off of the unconditioned stimulus of shock and a group of animals where the conditioned stimulus was turned off at the time of the unconditioned stimulus's onset.

The interpretation of these results is that the animals learned to *fear* the conditioned stimulus by virtue of its contiguity with the onset of pain. This anticipation of pain gave rise to trial-and-error behavior out of which was differentiated the running response, which was reinforced by fear reduction or the avoidance of pain. The acquisition of the fear reaction was not furthered, as reinforcement theory would predict, by having the warning signal overlap and end in contiguity with the reinforcing state of affairs provided by the termination of the shock.

The more crucial experiment, yet to be done, would involve the testing of the hypothesis that some autonomically mediated reaction, taken as an index of fear, will be attached to some conditioned stimulus by virtue of its association by contiguity with the onset of noxious stimulation, whereas it will not become attached any more effectively under conditions of reinforcement.

Experimentation with viscerovascular reactions presents many problems, however, and there is little in the literature that can be brought directly to bear on this issue. Indirect evidence is presented in the cited publications of Schlosberg and Skinner and is thoroughly reviewed by Mowrer (34).

While such interpretations are not crucial, much recent experimentation on secondary drives is readily assimilable into two-factor theory. Miller (29), for example, reports having trained rats by means of strong

⁵ The two-factor theories reviewed here may be contrasted with those of Stephens (57) and Maier and Schnierla (28). For a careful and trenchant critique of these points of view, see Kendler and Underwood (23).

shock to escape from a white compartment with a grid floor through an open door into a black compartment without a grid. Subsequently, the animals, without shock or noxious stimulation of any kind, learned a new habit—rotating a little wheel to open the door, which had been closed, in order to escape from the white compartment to the black one. This was interpreted to mean that the secondary drive of fear had been acquired and that its termination could be used as reinforcement for striped muscle responses. In terms of the two-factor formulation, the rats learned to run into the black box by virtue of the reinforcement provided by pain reduction. At the same time, however, fear or the visceral component of pain became conditioned to the cues of "whiteness and grid floor" associated with the *onset* of shock. The conditioned fear then served as the drive on the basis of which the wheel-rotating habit was learned without benefit of further primary drive arousal through shock.

It would seem, then, that in spite of its present tentative status, a two-factor theory of learning—holding that adaptive, striped muscle habits are built up according to the principle of reinforcement whereas anticipatory, "emotional" reactions, probably viscerovascular in nature and having drive properties, are acquired according to the principle of contiguity—has the greatest explanatory and predictive power at the moment.

LEARNING THEORY AND PSYCHOTHERAPY

How can such a conception of learning be applied to psychotherapy to cover the elements of the psychotherapeutic process common to all forms of counseling? It will be recalled that the problem of therapy is essentially that of somehow ridding the patient of neurotic anxiety, which supports his persistent non-integrative defenses and accounts in large measure for his "unhappiness." The tools used by all therapists to accomplish this job are those of conversational content and the therapeutic relationship.

Therapy as the acquisition of symbolic controls. Shaffer (49) suggests that psychotherapy be conceptualized in terms of the patient's acquisition of language symbols by which he can more effectively control his non-integrative behavior. The rationale of this approach is based on the observation that an outstanding characteristic of the maladjusted is their inability to control their own acts; in their own terms, "I know I should (or shouldn't) do this, but I just can't (or must)." Since "normal" people seem to control their behavior by means of symbols—including subvocal and gestural symbols—Shaffer's notion seems at first blush to follow readily.

Such an idea is also more or less explicit in Shaw's (50) analysis of repression and insight. He argues from Mowrer and Ullman's (38) point that,

The common denominator in all . . . forms of non-integrative behavior seems to be the inability to use symbols appropriately as a means of bringing remote as well as immediate consequences into the present in such a manner that they may exert an influence proportional to their objective importance (p. 81).

Shaw moves from here to the contention that therapy is a process by which non-integrative behavior is eliminated by the making available of symbols, holding that the symbols become cues for the more remote punishing consequences of neurotic defenses.

It is not quite clear, however, according to either Shaffer or Shaw, what the symbolization at which therapy aims might be. If it is the symbolizing of acts which have been repressed, there is no indication of how such a procedure would accomplish anything more than the release of a flood of anxiety heretofore held in check—albeit imperfectly—by the repression mechanism. On the other hand, if the symbols made available by therapy amount only to accurate predictions of the consequences of the client's non-integrative behavior, their utility is questionable on several grounds: First, most clinical patients are only too sharply aware of the self-defeating nature of their activity; their complaint is that they don't know why they engage in it and at the same time seem unable to avoid it. Second, some cases (especially those who have been formally psychoanalyzed) demonstrate a remarkable glibness—sometimes quite accurate—about their own defenses and yet are anxiety ridden on the one hand and socially somewhat obnoxious on the other. It is probably these instances which gave rise to H. M. Johnson's (20) rather oversevere recent strictures on psychoanalysis as therapy and as rationale. Third, there is a question as to whether or not simply making available symbols which can arouse at an earlier point in the temporal sequence the anxiety that accrues from future punishment amounts to anything more than a more effective punishment of the already non-integrative response. In this case, there may be the danger of the repression of one mechanism while another, equally self-defeating, is developed as a defense against a compounded neurotic anxiety, now attached not only to the ineffectively repressed impulses which existed prior to "therapy," but also to those incipient tendencies connected with the defense mechanism which has undergone the "punishment" of having its hurtful ultimate consequences symbolically brought into the psychological present. Thus, if a clinician is dealing with a patient whose anxiety has its origin in the faulty repression of aggressive tendencies and defends himself against it by social withdrawal, the anxiety may be compounded by making the damaging effects of the mechanism more apparent through the providing of symbols within the therapeutic context. All this is not to be construed as an attack on the Shaw-Shaffer hypothesis; as a matter of fact, it seems to describe quite adequately one segment of the therapeutic process. It is merely an effort to point out that such an hypothesis does not seem quite to account for *everything* that happens in psychotherapy.

A somewhat different suggestion, here proposed, is this: If neurotic anxiety is produced by the repression of some unextinguished response, it should follow that the anxiety can be dissipated in one of two ways—either by the elicitation of unreinforced occurrences of the response, thus leading to extinction, or by the connecting of a different affect to response tendencies which have undergone repression. With respect to the illustrative case mentioned above, anxiety could be dispelled either through eliciting self-initiated behavior and failing to reinforce it until extinction occurred, or through forming a bond between the tendencies to self-initiated behavior and some non-anxious visceral reaction which will supplant the connection between anxiety and the repressed activity. In either case, the Shaw-Shaffer notion holds as the first step in therapy, the bringing into communicability (consciousness) of the tendency that has undergone repression.

This lifting of repression is what is usually known as insight. When the patient is able to verbalize the repressed tendencies fundamentally associated with his anxiety, he "sees" or demonstrates insight. It is difficult to understand, however, why this should be equated with cure, regardless of how important it is as a step toward psychological recovery. Merely being able to talk about the cues for anxiety does not make them any less terrifying. Extinction or counter-conditioning is still necessary.

Whether the extinction or the counter-conditioning technique is preferable depends in part on the desirability of the repressed behavior. In the case of self-initiated activity, the question seems rather clear. Socialization has been defined (35) as the process of developing from a dependent infant into an independent and dependable adult. The extinction of tendencies toward self-initiated "responsible" behavior would mean the continuation of dependence and infantilism. It seems probable that few clinicians would look upon this as a suitable therapeutic goal. The same thing might well be said of most of the impulses which typically undergo repression, sexuality being a case in point. The frigid wife, raised under conditions of puritanical restrictiveness, might well find some immediate relief from anxiety by having her repressed sexual impulses extinguished (if this is possible); but it is doubtful that such a procedure would be helpful in her marriage.

The counter-conditioning hypothesis. The hypothesis of counter-conditioning is suggested as somewhat more tenable. It involves the following set of notions: The conversational content aspect of counseling consists in the symbolic reinstatement of the stimuli which produce and have produced the patient's anxiety. Through his words to the therapist, the client, on a symbolic level, again "lives through" the stimulus situations which were painful to him, in which he underwent punishment, and which initiated the repression sequence. This constitutes the lifting of repression, the introduction into communicability of the re-

pressed tendencies, the development of insight. This proceeds essentially by the therapist's reinforcing by his acceptance and his sympathetic participation of the patient's self-revelatory behavior. At the same time, the discussion of the client's anxiety is being carried on within the context of the unique patient-therapist relationship. This is conceived as an unconditioned stimulus for feelings of pleasure, acceptance, security—non-anxious affective reactions. The therapeutic process consists in the establishment of a bond between the symbolically reproduced stimuli which evoke and have evoked anxiety—chiefly the cues associated with the incipient movements toward performing some repressed activity—and the non-anxiety, i.e., comfort and confidence, reactions made to the counseling relationship.

Such a formulation goes somewhat beyond the bounds of "emotional" learning as accounted for by the two-factor theories briefly discussed above. They are chiefly concerned with the learning of fear or anxiety, basic secondary drives. While the idea presented here may be an extension of the theory that its protagonists would find unacceptable, there seems to be no reason why the principle of contiguity should not apply to viscerovascular reactions that are "pleasant" as well as to those which are "unpleasant"; as a matter of fact, such an application seems to be demanded if the learning of affects is governed by a single principle. The conceptualization proceeds in this wise: Affects possessing drive value—fear, anxiety, and anger⁶—are learned by virtue of the association by contiguity of the visceral aspects of some primary drive with concurrent external stimuli. The so-called "positive" or "pleasurable" affects are learned by virtue of the association by contiguity of proprioceptive cues set up at the onset of drive reduction with concurrent external stimuli. It is quite possible that Murray's (39) scheme for conceptualizing motivation in terms of goals is analyzable on some such basis as this latter notion.

Hull (19) seems to use a similar idea when he defines secondary reinforcement in terms of a stimulus situation which has been closely and consistently associated with the occurrence of need reduction. Experimental animals thus develop "needs" for poker chips, tones of given frequency, black compartments rather than white, etc. Likewise, the judgmental theory of affections, as proposed by Carr (5) and expanded upon and experimentally verified by Peters (41, 42), is fully consonant with the suggestion here proposed as fundamental in therapy. According to these writers, the pleasantness or unpleasantness of objects is a function of their association with "satisfying" or "unsatisfying" events in experience. Integrating this with the aspect of two-factor theory that deals with the learning of affects, "satisfying" events

⁶ The inclusion of anger in this list of secondary drives is somewhat cavalier. Virtually nothing is known of the conditions under which the learning of anger takes place, and it is certainly not assured that it derives from pain.

in experience are those correlated with drive reduction; "unsatisfying" events in experience are those correlated with drive onset.⁷

To return to the counter-conditioning hypothesis in psychotherapy, a rather striking analogy may be pointed out between this formulation and the now famous experiment of Mary Cover Jones (21) with the boy Peter. It will be remembered that Peter was a three-year-old with a number of acquired fears of various objects, including small white furry animals. In an effort to eliminate these fears, Dr. Jones attempted a counter-conditioning procedure. At lunch time, just as the child began to eat a meal which included his favorite dishes, a white rabbit was introduced in a wire cage at the end of the room, far enough away not to disturb the boy's eating. Each day the animal was brought a little closer until finally Peter could eat with one hand while stroking the rabbit with the other. Further tests showed that the newly conditioned "comfort" reaction to the rabbit had generalized to a large number of other, formerly fear-evoking stimuli such as rats, frogs, cotton, and fur rugs.

The meaning of these results is that a new connection was formed between the stimuli (rabbit) which produced a fear reaction and the comfort reaction made to the stimulus of the lunch with all its various cues. The necessary condition for the formation of this new connection was contiguity of the noxious stimulus and the comfort reaction aroused by the unconditioned luncheon stimulus situation. The problem of how to pair the stimuli so that those connected with the meal did not come to evoke fear does not affect the fundamental point of contiguity as the basis for the establishment of the new bond, but is merely a matter of the spatial and temporal patterning of stimuli common to most experimentation under the conditions of classical conditioning.

The main objection to this analogy probably rests on the point that Peter was troubled by a fear rather than an anxiety—that is, an affective reaction, uncomplicated by repression, made to external stimuli rather than to some impulse to behave in a tabooed way. The objection is certainly granted and actually implies the basis for the first step in therapy, the uncovering by use of the conversational content of therapeutic interviews of the repressed impulses. Before counter-conditioning can occur, the stimuli connected with anxiety must be brought into communicability, where they can be symbolically reinstated at the appropriate times. Insight is a prior condition of counter-conditioning.

A second objection that can be raised to the counter-conditioning notion is this: If therapy is simply a matter of connecting anxiety-provoking stimuli with some comfort reaction, why is it not therapeutically effective to think of one's troubles while lying in a comfortably warm tub?⁸ There seem to be three answers to this. First, to a degree it is effective. The widespread method of combatting the "blues" by

⁷ It is interesting to speculate as to whether or not this is the mechanism underlying the acquisition of aesthetic tastes, preferences, and other "likes" and "dislikes." The implications for a psychological approach to evaluative behavior are obvious.

⁸ This point was raised in a very helpful personal communication from Dr. John P. Seward. The replies offered to the objection, however, are not chargeable to him.

means of a shower is directly in point, as is the use of continuous baths and warm packs in mental hospitals. The real problem is: Why is such a procedure less effective than psychotherapy? This gives rise to the second answer, which is that thinking of one's troubles while lying in a comfortably warm tub is usually of little help in creating insight, symbolically re-introducing the relevant anxiety-producing stimuli. The bath is of little assistance in bringing forbidden impulses into communicability, hence the "therapeutic effects" of the bath are of short duration. The third reply to such an objection is based on the fact that neurotic anxiety is primarily social in its inception. Sullivan (59) insists that this "interpersonal induction of anxiety, and the exclusively interpersonal origin of every instance of its manifestations, is the unique characteristic of anxiety and of the congeries of more complex tensions . . . to which it contributes." This squares perfectly, of course, with the concept of repression and the role it plays in anxiety theory. If neurotic anxiety is an anticipation of punishment for the performance of some tabooed act, it follows that the taboo must have been laid down and enforced through some kind of social medium. Consequently, one would expect in the light of such social origins that the elimination of anxiety would be facilitated by the presence of certain social factors in therapy—provided in this case by the patient-therapist relationship.

This last point also bears on the function of catharsis in psychotherapy. It is a commonplace experience among clinicians to have clients say, after a period of vigorous abreaction, "I've thought about that a lot, but I've never said it to anybody before. I feel a bit better now." This poses something of a conceptual difficulty, since it is hard to understand how the expression of an affect should dissipate an affect unless the expression has some effect on the maintaining stimulus conditions. Such an environmental modification certainly does not occur in counseling; and yet catharsis in the social situation of therapy (and possibly in other social situations) seems to bring some relief, whereas catharsis subvocally or made without the presence of a therapist or therapist-surrogate apparently does not. According to the formulation here offered, *catharsis will be effective when it involves (a) the symbolic reinstatement of the repressed cues for anxiety (b) within the context of a warm, permissive, non-judgmental social relationship*. Under these conditions the situation is ripe for counter-conditioning to take place, whereby the patient learns to react non-anxiously to the original stimuli.

The counter-conditioning hypothesis likewise bears on the problems of technique inherent in the directive-non-directive controversy. This argument can perhaps be more profitably stated this way: How much and what can the therapist do to help reinstate symbolically the anxiety-arousing stimuli acting on the patient without endangering the relationship (i.e., weakening the relationship-comfort bond)? Asked in these terms, the question bears on the first step in counseling, that of lifting repressions or developing insight, and becomes the purely empirical

matter of determining the categories of counselor response that most effectively further the bringing into communicability of repressed impulses. On somewhat dangerous *a priori* grounds it would seem that interpretation, probing and other more active procedures would be useful unless introduced too preemptorily or too early into therapy, thereby destroying the patient-therapist relationship. That this occurs is not denied, but to attack such techniques as being of no value because they are sometimes misused seems somewhat absurd. The situation is analogous to bringing the rabbit too far into Peter's lunch room too early and connecting the fear reaction to the animal to the stimulus complex of food, room, high chair, and so forth. It seems somewhat nonsensical to argue that the baby should be thrown out with the bath water simply because it is still a bit grimy. One wonders if Peter would have overcome his fear of rabbits had he only been thoroughly "accepted" without ever having any help in reencountering the noxious stimulus in a secure and "pleasant" situation.

The directive-non-directive controversy may well reduce to a consideration of the types of case for which each is best suited. It can be hypothesized that more non-directive approaches will be more likely to succeed with those clients who have few and relatively unsevere repressions, some insight into the sources of their anxiety, and a capacity to relate easily to the therapist. These are cases which do not require much help in *discovering* the anxiety-producing stimuli; they do need assurance from a counselor that they may talk about them in his presence with complete impunity. Conversely, more interpretative methods by hypothesis will be of greater effectiveness with cases characterized by higher defenses, greater repression, and less initial insight. It must be emphasized, however, that all this is a matter of the empirical determination of what techniques work best for given cases so far as the lifting of repressions is concerned. The hypothesis of counter-conditioning is still the means of explaining the diminution of anxiety after insight has been developed.

If this formulation is correct, how can various failures of counter-conditioning methods in psychological treatment be answered? Voegtlin's (63) work with alcoholics is typical. This clinician attempted to cure his patients of drinking by having them take whiskey so heavily dosed with a powerful emetic that vomiting to the point of pain was immediately induced. Results were disappointing. Most of his cases did not build up more than momentary conditioned aversions to alcohol. Of those few who became conditioned against liquor over a period of time, several showed symptom substitutions, e.g., the development of psychosomatic symptoms or neurotic syndromes instead of alcohol addiction.

The first objection to such a procedure is that it consists in a direct attack on the symptomatic mechanism rather than on the underlying anxiety. If the anxiety reduction occurring from drinking were greater

than the pain of the treatment, the treatment would have very nearly as little effect as strongly advising the patient "to get on the wagon." The ineffectiveness of "hangovers" is relevant in this connection. Second, if the alcohol addiction were wiped out by virtue of the conditioning procedure, the underlying anxiety would be unaffected, and one would therefore expect that the patient would develop some other persistently non-integrative way of reducing it. Third, the treatment situation contains too many elements of attempting to eliminate a response by merely punishing it. The inefficacy of such methods has already been discussed. Thus, an objection based on such therapeutic experience fails to carry much weight.

Reeducation in psychotherapy. Does the point of view developed here overlook this notion in the therapeutic armamentarium? On the contrary, it fully includes it as an important third aspect of counseling, along with the lifting of repression and the counter-conditioning of anxiety. Following the development of insight, as anxiety is dissipated through conditioning, the patient typically begins to plan. His first tentative steps in this direction may take the form of asking, "What shall I do?" Or it may be a more vigorous exploration of the possible consequences of projected steps. Here the therapist may be of assistance in helping his client to formulate goals clearly and to consider realistically the various behavioral methods he might employ to reach them. This constitutes a law-of-effect learning situation in which reinforcement is produced through the patient's own verbal self-approval or self-disapproval, based in part on the predictions of consequences which the counselor can help him arrive at. In a sense, this constitutes the "rational" exercise of symbolically mediated self-control of which Shaw and Shaffer may be speaking. It is rational insofar as the behavior selected is founded on some consideration of its probable remote outcomes rather than on its immediate value as an anxiety-reducing agent, and it is "responsible" insofar as it is chosen⁹ in terms of the patient's own values as of the moment of choice. The counselor does not direct; he merely helps the client work out relatively accurate estimates of the consequences. If a particular behavior pattern is rejected, it merely undergoes a voluntary suppression or is extinguished through failure of reinforcement without being forced into incommunicability and becoming a stimulus for anxiety, as is the case in the repression of punished tendencies. Through this symbolic trial and error, then, the patient develops, according to the principle of reinforcement, a tentative plan of integrative behavior based on rational considerations to supplant his former pattern of persistent non-integrative behavior based on the immediate necessity of reducing anxiety regardless of the ultimate cost.

⁹ Lest the language used here seem flavored too heavily with free will, reference is made to Hall's (16) paper, in which the problem of choice within a deterministic philosophy is discussed.

SUMMARY

A learning theory interpretation of psychotherapy must take into account (a) the fact that all forms of psychotherapy are able to claim cures, (b) the similarity of clinical cases in terms of neurotic anxiety and its defenses, (c) the common goal of psychotherapies of the diminution of anxiety, and (d) the fact that all clinicians employ as their chief techniques conversational content and the therapeutic relationship.

It is here proposed that psychotherapy occurs through three inter-related processes: first, the lifting of repression and development of insight through the symbolic reinstating of the stimuli for anxiety; second, the diminution of anxiety by counter-conditioning through the attachment of the stimuli for anxiety to the comfort reaction made to the therapeutic relationship; and third, the process of reeducation through the therapist's helping the patient to formulate rational goals and behavioral methods for attaining them.

Such a scheme seems to harmonize most effectively with a two-factor learning theory of the type most recently developed by Mowrer (34). Such a theory conceives of skeletal muscle responses as being acquired through the principle of reinforcement, whereas viscero-vascular, "emotional" reactions are acquired according to the principle of contiguity.

This formulation is certainly not to be regarded as anything final. It leans rather too much on plausible but inadequately tested hypotheses and on scientifically tenuous analogies. It is offered only as a preliminary attempt to effect a *rapprochement* between psychotherapy and general psychology, and to organize some of the phenomena of clinical practice within the framework of systematic behavior theory.

BIBLIOGRAPHY

1. ALEXANDER, F., & FRENCH, T. *Psychoanalytic therapy*. New York: Ronald Press, 1946.
2. ALLEN, F. *Psychotherapy with children*. New York: W. W. Norton, 1942.
3. AXLINE, V. *Play therapy*. Boston: Houghton Mifflin, 1947.
4. CAMERON, N. *The psychology of the behavior disorders*. Boston: Houghton Mifflin, 1947.
5. CARR, H. *Psychology*. New York: Longmans, Green, 1925.
6. COMBS, A. W. Phenomenological concepts in nondirective therapy. *J. consult. Psychol.*, 1948, 12, 197-208.
7. DARLEY, J. Review of "Counseling and psychotherapy." *J. abnorm. soc. Psychol.*, 1943, 38, 199-201.
8. DEJERINE, J., & GAUCKLER, E. *The psychoneuroses and their treatment by psychotherapy*. Philadelphia: J. B. Lippincott, 1913.
9. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, No. 3.
10. FINESINGER, J. E. Psychiatric interviewing. *Amer. J. Psychiat.*, 1948, 105, 187-195.
11. FRENCH, T. Interrelations between psychoanalysis and the experimental work of Pavlov. *Amer. J. Psychiat.*, 1933, 12, 1165-1203.

12. FREUD, S. *New introductory lectures on psychoanalysis*. New York: W. W. Norton, 1933.
13. FREUD, S. *A general introduction to psychoanalysis*. New York: Live-right, 1935.
14. FREUD, S. *The problem of anxiety*. New York: W. W. Norton, 1936.
15. GUTHRIE, E. R. A theory of learning in terms of stimulus, response, and association. *National Society for the Study of Education*, 41st Yearbook. Bloomington: Public School Publ. Co., 1942. Pp. 17-60.
16. HALL, E. W. An ethics for today. *Amer. J. econ. Sociol.*, 1943, 2, 444-446.
17. HORNEY, K. *The neurotic personality of our time*. New York: W. W. Norton, 1937.
18. HULL, C. A functional interpretation of the conditioned reflex. *Psychol. Rev.*, 1929, 36, 498-511.
19. HULL, C. *Principles of behavior*. New York: Appleton-Century, 1943.
20. JOHNSON, H. M. Psychoanalytic therapy versus psychoanalytic rationale. *Amer. Psychologist*, 1948, 3, 337. (Abstract.)
21. JONES, M. C. A laboratory study of fear: the case of Peter. *Ped. Sem.*, 1924, 31, 308-315.
22. KENDLER, H. H., & MENCHER, H. C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, 38, 82-88.
23. KENDLER, H. H., & UNDERWOOD, B. J. The role of reward in conditioning theory. *Psychol. Rev.*, 1948, 55, 209-215.
24. KRAINES, S. *Treatment of the neuroses and psychoses*. (2nd ed.) Philadelphia: Lea & Feabiger, 1943.
25. KUBIE, L. S. Relation of the conditioned reflex to psychoanalytic technique. *Arch. Neurol. Psychiat.*, 1934, 32, 1137-1142.
26. LOUCKS, R. B. The experimental delimitation of neural structures necessary for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1935, 1, 5-44.
27. LOUCKS, R. B., & GANTT, W. H. The conditioning of striped muscle responses based on faradic stimulation of dorsal roots and dorsal columns of the spinal cord. *J. comp. Psychol.*, 1938, 25, 415-426.
28. MAIER, N. R. F., & SCHNIERLA, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, 49, 117-134.
29. MILLER, N. E. Studies of fear as an acquirable drive. I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, 38, 89-101.
30. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
31. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univ. Press, 1941.
32. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, 46, 553-565.
33. MOWRER, O. H. The law of effect and ego psychology. *Psychol. Rev.*, 1946, 53, 321-334.
34. MOWRER, O. H. On the dual nature of learning—A reinterpretation of "conditioning" and "problem-solving." *Harvard educ. Rev.*, 1947, 17, 102-148.
35. MOWRER, O. H., & KLUCKHOHN, C. A dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1943. Pp. 69-135.
36. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, 39, 29-50.
37. MOWRER, O. H., & SUTER, J. Further evidence for a two-factor theory of learning. Unpublished study.

38. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, 52, 61-90.
39. MURRAY, H. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
40. O'CONNOR, F. J. Recency or effect? A critical analysis of Guthrie's theory of learning. *Harvard educ. Rev.*, 1946, 16, 194-206.
41. PETERS, H. N. The judgmental theory of pleasantness and unpleasantness. *Psychol. Rev.*, 1935, 42, 354-386.
42. PETERS, H. N. Experimental studies of the judgmental theory of feeling: I. Learning of positive and negative reactions as a determinant of affective judgments. *J. exp. Psychol.*, 1938, 23, 1-25.
43. RAZRAN, G. S. The law of effect or the law of qualitative conditioning? *Psychol. Rev.*, 1939, 46, 445-463.
44. ROGERS, C. *Counseling and psychotherapy*. Boston: Houghton Mifflin Co., 1942.
45. ROGERS, C. Some observations on the organization of personality. *Amer. Psychologist*, 1947, 2, 358-368.
46. ROSENZWEIG, S. Some implicit common factors in diverse methods of psychotherapy. *Amer. J. Orthopsychiat.*, 1936, 6, 412-415.
47. SANFORD, R. N. Psychotherapy and counseling: Introduction. *J. consult. Psychol.*, 1948, 12, 65-67.
48. SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, 44, 379-394.
49. SHAFFER, L. The problem of psychotherapy. *Amer. Psychologist*, 1947, 2, 459-467.
50. SHAW, F. A stimulus-response analysis of repression and insight in psychotherapy. *Psychol. Rev.*, 1946, 53, 36-42.
51. SHOBEN, E. J., JR. A learning-theory interpretation of psychotherapy. *Harvard educ. Rev.*, 1948, 18, 129-145.
52. SKINNER, B. F. Two types of conditioned reflex and a pseudo-type. *J. gen. Psychol.*, 1935, 12, 66-77.
53. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
54. SNYDER, W. U. An investigation of the nature of non-directive counseling. *J. gen. Psychol.*, 1945, 33, 193-224.
55. SPENCE, K. W., & KENDLER, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, 38, 106-109.
56. SPENCE, K. W., & LIPPITT, R. An experimental test of the sign-gestalt theory of trial-and-error learning. *J. exp. Psychol.*, 1946, 36, 491-502.
57. STEPHENS, J. M. Expectancy vs. effect-substitution as a general principle of reinforcement. *Psychol. Rev.*, 1942, 49, 102-116.
58. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington: The William Alanson White Memorial Foundation, 1947.
59. SULLIVAN, H. S. The meaning of anxiety in psychiatry and in life. *Psychiatry*, 1948, 11, 1-13.
60. TAFT, JESSIE. *The dynamics of therapy*. New York: Macmillan, 1933.
61. TOLMAN, E. C. The determiners of behavior at the choice point. *Psychol. Rev.*, 1938, 45, 1-41.
62. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, 55, 189-208.
63. VOEGTLIN, W. L. The treatment of alcoholism by establishing a conditioned reflex. *Amer. J. med. Sci.*, 1940, 109, 102.
64. WHITE, R. *The abnormal personality*. New York: Ronald Press, 1948.
65. WHITING, J. W. M. *Becoming a Kwoma*. New Haven: Yale Univ. Press, 1941.
66. WILLIAMSON, E. G. *How to counsel students*. New York: McGraw-Hill, 1939.

Received March 7, 1949.

STATISTICAL METHODS APPLIED TO RORSCHACH SCORES: A REVIEW¹

LEE J. CRONBACH

Bureau of Research and Service, University of Illinois

While the Rorschach test grew out of clinical investigations, and is still primarily a method of individual diagnosis, there is increasing emphasis on statistical studies of groups of cases. On the whole, the statistical methods employed have been conventional, even though the Rorschach test departs in many ways from usual test methodology. The present review proposes to examine the methods which have been employed to deal with Rorschach data, and to evaluate the adequacy of those often used. It attempts to provide a guide to future investigations by indicating statistically-correct studies which can serve as models. There is no intent here to review the generalizations about the test arising from these studies, or to call into question general research procedures, sampling, and other aspects of the studies.

This report may be considered an extension of a review by Munroe (41). In 1945, she considered the objectivity of previous Rorschach research. She distinguished between the goals attainable by clinical intuitive interpretation and the goals to be reached by more quantitative procedures. She traced the trend in Rorschach literature, noting the gradual decrease in studies based solely on impressionistic treatment of data or on mere counting of scores, and the introduction of significance tests, standard deviations, and other signs of adequate effort to test generalizations statistically. She also pointed out some errors in statistical thinking that lead to faulty conclusions about the Rorschach test. Munroe takes the position, and the writer fully concurs, that statistical research on the Rorschach test is not only justifiable, but indispensable. The flexibility of clinical thinking creates excellent hypotheses, but these hypotheses can only be established as true by controlled studies. Among the propositions suggested by clinical work, some are certainly untrue, due to faulty observation, inadequate sampling, and errors of thinking. Statistical controls are essential to verify theories of test interpretation, and to validate proposed applications of the test. Even though the clinician studying one person makes no use of statistics, he employs generalizations about the test which must rest on scientifically-gathered evidence. Munroe demonstrated that the Rorschach test lends itself to objective studies; the writer reviews the same ma-

¹ The writer wishes to express appreciation to Frederick Mosteller and to N. L. Gage, who read this manuscript and contributed many suggestions for its improvement.

terial more technically to evaluate the soundness of the statistical procedures on which the conclusions are based.

CLINICAL TREATMENTS OF DATA

While this paper deals principally with statistical methods applied to *raw* Rorschach data, we shall consider briefly the statistical procedures used when clinically interpreted case records are used in a study. The Rorschach record is usually interpreted qualitatively and in a highly complex manner when the test is given in the clinic, and many studies have been based on these interpreted records. In only a few studies of this type do statistical problems arise.

Dichotomized Rorschach ratings. In one type of study, the interpreter of the records makes a final summary judgment, dividing the records into such groups as "adjusted-maladjusted" or "promising-unpromising," etc. This method is most used for validation studies, where the Rorschach judgment is compared with a criterion of performance or with a judgment from some other test. Simple statistical tests suffice to test the degree of relationship. If the criterion is expressed in two categories (as when the criterion indicates success or failure for each case), chi-square is simple and appropriate. This is exemplified in a study of success of Canadian Army officers (51), where a prediction from the Rorschach is compared with a later rating of success and failure. If the criterion is a set of scores on a continuous scale, bi-serial r is usually an adequate procedure. In bi-serial r , one assumes that the dichotomy represents a continuous trait which is normally distributed. This assumption is generally acceptable for personality traits and for ratings of success.

Rorschach ratings on continuous scale. In some studies, the Rorschach interpretation is reported in the form of a rating along a scale, rather than as a dichotomy. When the criterion is dichotomous, bi-serial r is appropriate. (E.g., a prediction of probable pilot success is so correlated with elimination-graduation from training, 21, p. 632.) For a continuous criterion, like grade-average, product-moment r is conventionally used.

These methods are not entirely satisfactory, because of a limitation of rating scales. If units on the rating scale are not psychologically equal, the correlation may not indicate the full size of the relationship. If ratings are careful, one can assume that men rated "Good" are superior to men rated "Fair," and that men rated "Excellent" are superior to both of these. But it may be unwise to assume that the jump from "Good" to "Excellent" is equal to the jump from "Fair" to "Good," as one automatically does in correlating. One solution to this difficulty is to assume that the trait rated is normally distributed in the men studied. Then we can condense the five-point scale into a dichotomy, which

is the case discussed in the preceding paragraph. Alternatively, one may convert the ratings into scaled values which will yield a normal distribution (34). Bi-serial r is then appropriate, if the criterion is dichotomous. Similar reasoning applies to the correlation of a rating with a continuous criterion; one will obtain the most meaningful results by dichotomizing the rating and using bi-serial r , or by normalizing before using product-moment r . These suggestions are summarized in Table I.

TABLE I
PREFERRED METHODS FOR COMPARING RORSCHACH INTERPRETATIONS WITH
CRITERIA OF VARIOUS TYPES

Criterion	Judgment made from Rorschach	
	Dichotomy	Continuous scale, unequal units
Dichotomy	χ^2	χ^2 after dichotomizing rating; r_{bis} after normalizing rating*
Continuous scale, unequal units	χ^2 after dichotomizing criterion; r_{bis} after normalizing criterion*	χ^2 after dichotomizing both variables; r_{bis} after normalizing one, dichotomizing the other; product-moment r after normalizing both
Continuous scale, equal units	r_{bis} *	r_{bis} after dichotomizing rating; product-moment r after normalizing rating

* Point bi-serial must be used if the two parts of the dichotomy cannot reasonably be considered subdivisions of a continuous scale.

Munroe (42), comparing a Rorschach adjustment rating with success in academic work, where both variables were reported on a four-category scale, used a coefficient of contingency. Where the correlation surface is nearly normal, this coefficient with proper corrections should give approximately the same result as the product-moment r for normalized data, corrected for broad categories. Yates (70) has recently offered an alternative method of adapting the contingency method to take advantage of trends in the relationship between variables expressed as ordered categories.

Matching methods. Another favorite technique for evaluating Rorschach results is blind-matching, which permits a study of each case "as a whole." When a set of Rorschach records (interpreted or not) and another set of data regarding the same individuals are available, one may request judges to match the two sets in pairs. The success of

matching is evaluated by a formula developed by Vernon (66). An example of its use is a study by Troup (62), in which judges tried to match two Rorschach records for each person. One hundred fourteen matches were correct out of a possible 120, judges considering five pairs at a time. By the Vernon formula, this corresponds to a contingency coefficient of .88. A coefficient of .40 was obtained when judges attempted to match the record of each case with that of his identical twin. Another excellent illustration of the method is provided by J. I. Krugman (31), who used it to establish that different evaluations of the same Rorschach protocol could be matched, and that the interpretations could be matched to the raw records and to criteria based on a case-study.

The limitations of this method are not statistical; they lie more in the human limitations of judges. A portrait based on the Rorschach may be nearly right, yet be mismatched because of minor false elements. Matching, on the other hand, might be excellent, even perfect; the study would still not guarantee that each element in each portrait was correct, especially if the subjects were quite different from each other. In fact, the portrait might be seriously wrong in some respects, without preventing matching.

A complex modification of the blind-matching method has been proposed and tried by Cronbach (9). Judges are asked to decide whether each statement on a list fits or does not fit a case described in a criterion sketch. Since only about one-third of the statements in the list were actually made about the given case, one can test by chi-square whether the matching is better than chance. (The method yields many interesting types of information: (a) an all-over estimate of the validity of predictions with relation to the criterion, (b) a separate estimate of the validity of the description for each case or for subgroups, and (c) an estimate of the validity of statements dealing with any one aspect of personality (e.g., social relations).)

ERRORS IN STATISTICAL STUDIES

The majority of statistical studies with the Rorschach test have treated Rorschach scores directly, with clinical judgment eliminated. This is an important type of investigation, which presents numerous problems. Before considering general questions of procedure, however, it is necessary to deal with several errors and unsound practices found in the literature reviewed. These miscellaneous errors must be pointed out lest they be copied by later investigators, and to suggest that the studies in which the errors occurred need to be reevaluated.

Significance tests for small samples. The critical ratio is not entirely satisfactory when applied to small samples. When there are fewer than 30 cases per group, the *t* test is preferable. This would apply, for exam-

ple, in Goldfarb's (19) comparison of obsessionals with supposedly normal adolescents. His significance ratios are a bit too high, since he used the formula $\text{diff.}/\sigma_{\text{diff.}}$ with groups of 20 cases. (It may be noted also that Goldfarb's study does not permit sound generalizations about obsessionals as compared to other adolescents. The obsessionals had a mean IQ of 120 compared to 97 for the normals, so that differences between the groups may be due to intelligence rather than obsessional trends.)

Chi-square is generally useful for small samples, but it is important to apply corrections when the number of cases is below 50. This is especially important when the expected frequency in any cell of a 2×2 table is five or lower, under the null hypothesis. Many Rorschach studies fail to recognize the need for corrections, Kaback's (29, pp. 24, 38-39) being a striking example. She compares the distribution of such a score as M in each of two groups. To do so, she makes the distribution in a great number of intervals, with only a few cases per interval, and tests the similarity of the distributions by chi-square. In such a case, with many small cell frequencies, no significant result could be expected. Nor is it useful to inquire, as her procedure does, whether the precise distribution of M scores is the same for the two groups (in her case, pharmacists and accountants.) Her major question was whether one group used M more than the other, and this could be answered by dichotomizing the distribution and then applying chi-square, with proper correction. In applying chi-square to the 2×2 tables, one should as a standard practice apply Yates' correction (56, p. 169). The importance of this correction will be demonstrated in Table IV. (Where groups are dichotomized, it is best to make cuts toward the center, so that marginal totals will remain reasonably large. Special problems in the application of chi-square to successive tests of the same hypothesis, and to problems of goodness of fit, are discussed by Cochran (6).)

Tests for significance of difference in proportions. Throughout the Rorschach literature, the formula for the significance of differences between proportions is misused. The resulting inaccuracy is slight in most problems, fortunately. This error is common in other work, and even some statistics books appear to endorse the faulty procedure. The usual formula,

$$\sigma_{p_1-p_2} = \sqrt{\frac{pq}{N_1} + \frac{pq}{N_2}}$$

may not be entered with p_1 and p_2 , the proportions obtained in the two samples. Instead, one should substitute p_0 for p , where

$$p_0 = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}.$$

A significance test inquires whether p_1 and p_2 might arise by chance in sampling from a homogenous population in which the true proportion is p_0 (see 35, pp. 126-129). Employing p_1 and p_2 , instead of entering p_0 in both terms, almost always increases the critical ratio over what it should be. Because no correct model is found in the Rorschach literature, the following example is given using Hertz' data (25).

Five boys out of 41, and 0 girls out of 35 gave zero color responses.

$$p_0 = \frac{5 + 0}{76} = .066$$

$$s.d._{diff.} = \sqrt{\frac{.066 \times .934}{51.41} + \frac{.066 \times .934}{35}} = .057$$

$$p_1 = .122; p_2 = .00; \frac{\text{diff.}}{s.d._{diff.}} = \frac{.122}{.057} = 2.14 (P = .032)$$

This compares to the critical ratio of 2.41 ($P = .016$) computed by the formula Hertz and other workers have inadvisedly used.

The above computation is equivalent to the determination of significance by chi-square, and yields an identical result. But in this instance the expected frequencies are so low that the correction for continuity becomes important. Applying Yates' correction, we find that P becomes .10, and the reported difference is not significant.

Several studies use the formula for proportions in independent samples when the formula for paired samples should be used. Thus Hertz (25), to compare the 12-year-old and 15-year-old records of the same cases, should use a formula for correlated samples as given by Peatman (44, p. 407) or by McNemar (37; see also 13, 59). The correct formula would have yielded significant differences where Hertz found none. Other studies employing matched samples, where the significance of differences was underestimated by a formula for independent groups, are those of Hertzman and Margulies (27), Meltzer (39), M. Krugman (32), Richardson (48), and Goldfarb (20). In studies where the subjects were children varying widely in age, the proper formula would probably have yielded quite different results.

A study by Brown (4) committed this error and one even more serious. He compared records of 22 subjects without morphine and then with morphine. He found that 14 increased in R and 7 decreased. He then treated these as independent proportions of the 22 subjects, computing the critical ratio for the difference 64% minus 32%. These are not proportions in independent samples, and Brown's statistical tests are meaningless. No manipulation of the increase-decrease frequencies is as satisfactory for this problem as the formula given by McNemar. Brown could properly have set a cutting score (e.g., 20R) and compared the percentage exceeding this level with and without morphine.

Siegel's procedure (55), in which the "percentage incidence" of a factor in

one group is divided by the incidence in the second group, will be likely to produce misleading results.

An alternative formula for the significance of differences in matched groups is used by Gann (18). In applying the formula, however, a serious error was made. The formula given by Engelhart which Gann adopted is

$$\sigma_{\text{diff.}}^2 = (\sigma_{M_1}^2 - \sigma_{M_2}^2)(1 - r_{ij}^2)$$

r_{ij} is the correlation of the matching variables with the variable in which a difference is being tested. This formula may be extended to differences in proportions, although the estimated population value (p_0) for the proportion should be substituted for M_1 and M_2 , as explained above. Gann's major error was to use a value of .9741 for r_{ij} in all her calculations. From the context, this seems to be a multiple correlation of all matching variables with *all* dependent variables. The proper procedure, for any single significance test such as the proportion of cases emphasizing W , would be to correlate the matching variables with W -tendency alone. This correlation would almost certainly be close to zero. By the procedure Gann used, the critical ratios are very much larger than they should be. In one comparison where Gann reported a CR of 6.02 the writer has established that the true CR cannot be greater than 2.23, and is almost certainly less.

Comparisons of total number of responses. It is thoroughly unsound to compare the total number of responses of a given type in two samples. Swift (58) tested 37 boys and 45 girls. The boys gave a total of 248 F responses; all girls combined gave 246. Swift used chi-square, demonstrating that these 494 responses were divided in a way which departs significantly from the theoretical ratio 37:45. But this assumes 494 independent events in her sample whereas she really had 82. The F responses are not independent, since some were made by the same person. She might properly have used the t -test, applied to the means of the groups. The only correct way to use chi-square on her problem is to compare the number of cases exceeding a certain F score (cases, not responses, being the basis of sampling). A similar error has been made by Hertzman (26), Rickers (49, p. 231), and Werner (68).

Richardson (48) followed a different erroneous procedure. In her Table 9, she determined what proportion of all responses in each of her groups were W responses, and tested the difference in proportions for significance using the number of subjects in the denominator of the significance formula. The "proportion" she was studying is actually the ratio *Mean W/Mean R*, and the standard deviation of this is not correctly given by the formula $\sqrt{pq/N}$. If she must test the W/R ratio, in spite of the difficulties to be considered later, it is necessary to determine the ratio for each person separately and test differences between the groups in one of the conventional ways (e.g., chi-square, t -test, etc.).

Inflation of probabilities. Rorschach studies are peculiarly prone to an error which can arise in any statistical work. If a particular critical ratio or chi-square or t -test corresponds to a P of .05, we conventionally interpret that as statistically significant because "such a value would

arise by chance only once in twenty times." While this usually refers to once-in-twenty-samples, it may also be thought of as "once in twenty significance tests," if the several tests are independent. In some Rorschach studies, a vast number of significance tests are computed. Thus Hertz in one study reported the astonishing total of eight hundred significance tests (25). Many of these comparisons reach the one percent level or the five percent level, but even these are not all statistically significant. Quite a few of these differences did arise by chance, and unfortunately we cannot estimate how many because the tests were not experimentally independent. The proper procedure, in such a case, is to recognize that an inflation of P values has taken place. The analogy to monetary inflation is a fair one: The increase in the number of significance tests in circulation causes each P to have less worth than it would normally. We may accordingly raise our "price" arbitrarily, and insist that P reach a higher level than .05 before we label it "significant," and a higher level than .01 before we label it "very significant." Of the differences reported in the Rorschach literature as "significant at the 5% level," probably the majority are due to chance.

There are several ways in which significance levels may be inflated so that they become falsely encouraging. (One is the common procedure of testing differences on a great many Rorschach scores. This is of course sound practice, but one must then take the total number of significance tests into account in evaluating P .) The inflation is more subtle when the investigator rejects a large number of hypotheses by inspection without computing significance tests, and reports only a few significance tests. Thus Piotrowski and others (46) compared superior and inferior mechanical workers on "all the components used in conventional scoring as well as many others." They finally invented four composite scoring signs on which differences between the two samples were large enough to encourage a significance test. Suppose, for simplicity, that those four tests had yielded P 's of .02. The significance of those P 's must be minimized in view of the fact that four such differences were found in several hundred implied comparisons which were not actually computed, and two per hundred is chance expectation.

A comparable inflation arises when an investigator slices a distribution in order to take advantage of chance fluctuations and find some "hole" where a test will yield a low P . Hertz applied the formula for significance of the difference in proportions, to compare two groups on $M\%$ (Table II). She introduced a spurious element by slicing the $M\%$ distribution in so many places, and making so many significance tests. If a distribution is dichotomized in many ways, the chances of a "significant" difference rise greatly. Here only one test yielded a P of .05, out of nine attempted. The interpretation "It may be said with certainty, that more girls than boys at 15 years give over 11% M " (25, p. 180) is unjustified. In another sample this fluctuation would not occur. (It is not necessary to test explicitly all possible dichotomies for this type of

error to arise. If the investigator examines his distribution and makes his cut at the place where the difference is greatest, he has by implication examined and discarded all other possible hypotheses. One of the several studies where this occurs is that of Margulies, discussed later.)

Multiple correlation procedures give rise to a similar error. Suppose ten scores are tried as predictors. These scores might be combined in a prediction formula in an infinite number of ways. When an investigator

TABLE II
SIGNIFICANCE DATA REPORTED BY HERTZ FOR DIFFERENCES IN $M\%$ BETWEEN
15-YEAR-OLD BOYS AND GIRLS (25)

<i>Difference Tested</i>	<i>Critical Ratio</i>	<i>P</i>
Difference in means	1.47	.15
Difference in medians	2.32	.05
Difference in proportions		
in interval 0-1	.81	..
in interval 0-3	.81	..
in interval 0-5	1.83	.10
in interval 0-7	1.68	.10
in interval 0-9	.90	..
in interval 0-11	2.34	.05
in interval 0-13	1.24	..
in interval 0-15	1.81	.10
in interval 0-17	1.23	..

computes correlations and works out the best possible predictive combination for his particular data, he implicitly discards all the other combinations. Even though his combination gives a substantial multiple R for the original sample, it is certain to give a lower correlation in a new sample where the formula can no longer take advantage of chance fluctuations. The common practice of comparing two groups on a large number of signs and developing a checklist score in which the person is allowed one point for every sign on which the two groups differ, is open to the same objection. In a new sample many of these signs will no longer discriminate.² When a significance test is applied to a difference in checklist scores or to a multiple correlation in the sample on which the combining formula was derived, the significance test has only negative meaning. (If, even after taking advantage of chance differences, one's formula cannot discriminate, it is indeed worthless. But if the result gives a P better than .05, the formula may still be of no value.

² Harris (24) claims that in his experience the Rorschach behaves differently from other tests, and that signs found to differentiate in one sample are usually confirmed in other samples. This appears improbable on logical grounds, and no evidence in the literature supports such a statement.

Rorschach studies which have reported "significant" differences based on an empirical formula) without confirming them on fresh samples are those of Montalto (40), Harris and Christiansen (23), Hertzman, Or-lansky, and Seitz (28), and Ross and Ross (52). Thompson (60) reports spurious r 's but does not claim significance for them. Buhler and Le-fever (5, Tables X, XX) mix new cases with the sample used in deriving scoring weights, and therefore fail to provide an adequate test of signif-icance. Significance tests on fresh samples have been properly made by Guilford (21), Gustav (22), Margulies (38), Ross (50), and Kurtz (33). The latter gives a particularly clear discussion of the issue involved. In most studies, correlations nearly vanish when a Rorschach prediction formula is tried on a new sample.

(Still another method of inflating probabilities is to recombine groups of subjects in a way to maximize differences.) If one has several types of patients, all of whom earn different mean M scores, these groups may be recombined in many ways, and in one of the possible regroupings a pseudo-significant difference may be found. Rapaport and his coworkers (47) have carried inflation to bizarre levels. Not only did they consider scores in great profusion and in numerous combinations. They re-combined their subjects so that the number of implicit significance tests in their volume is incalculable. They began with subjects in 22 sub-groups. Significance tests were then made, on any score, between any pair of subgroups or combinations of them which seemed promising *after* inspection of the data. There were 231 possible pairs of subgroups, and an endless variety of combinations. Thus at times Unclassified Schizophrenics Acute were lumped with one, two, or more of the follow-ing: Paranoid Schiz. Acute, Simple Schiz., Uncl. Schiz. Chronic, Par. Schiz. Chr., Uncl. Schiz. Deteriorated; or with all the schizophrenics and preschizophrenics; or with Paranoid Condition, Coarctated Pre-schiz., Overrideational Preschiz., and Obsessive-Compulsive Neurosis. Such willingness to test any hypothesis whatever leaves these workers open to the charge of having regrouped their cases to augment differ-ences. They have undoubtedly reported differences which were created by artificial combinations of chance variations between groups. (Every time cases are recombined for a significance test, one must recognize that a large number of implied significance tests were also made, since many other recombinations were rejected without actual computation.)

Rorschach studies, because of the great number of scores and the large number of subgroups of subjects involved, are more prone to infla-tion than other research. The suggestions to be made for sound practice are these:

- ✓ 1. Compare the number of significant differences to the total number of comparisons in the study, both those computed and those rejected by implica-tion.

- ✓2. Raise the P value required for significance as the number of comparisons increases.
- ✓3. Never accept an empirical composite score or regression formula until its discriminating power has been verified on a new sample.
- ✓4. In general, do not trust significance tests unless the hypothesis tested was set up independent of the fluctuations of a particular sample.

These suggestions require that the investigator have clearly in mind the number of comparisons considered. Comparisons are of three types: those rejected as improbable before the data are looked at, i.e., before the study is begun; those not computed because a cursory inspection showed no apparent difference; and those computed. Sometimes the investigator begins with, say, five groups of subjects and ten scores, and frankly wants to unearth all possible differences between types of subjects. Then there are ten ways the groups may be paired against each other, and since each pair may be compared on each score, there are a total of one hundred comparisons in the study. If, on the other hand, the investigator sets out to check only certain relationships—"Schizophrenics differ from neurotics in $F + \%$," "Manics differ from all other groups combined in $FC:CF + C$ "—those limited hypotheses may be laid down in advance of the study, and only those comparisons are counted as implied significance tests. To avoid confusion, it is also well for the investigator to specify his cutting point, if a variable is to be dichotomized, before examining the differences between groups. This may be set by an arbitrary rule, for instance that each distribution is to be divided as near to its median as possible, or by an *a priori* decision to divide at some point such as $2M$. In essence, the investigator must ask himself before he gathers his data, "How many comparisons do I intend to look at, and charge myself for?" A P of .01 may be called significant if it is one of three comparisons charged for, but not if the investigator has looked at three hundred comparisons in order to salvage this one impressive value."

METHODS OF COMPARING GROUPS ON RORSCHACH SCORES

Necessity for Choosing between Statistical Procedures

Because Rorschach scores are numbers which can be added, averaged, distributed, etc., most investigators have used conventional mental-test statistics without question. The most common need for statistics is to compare the test scores of groups and determine the significance of differences. The prominent methods encountered in Rorschach literature are as follows: significance of difference between means (critical ratio or t -test); analysis of variance; bi-serial r ; significance of difference in proportions exceeding a particular score, or chi-square; and significance of difference between medians. }

Apart from such errors as those listed in the preceding section, there is no reason for considering any of the procedures under discussion as mathematically incorrect. If a significant difference is revealed by any proper significance test, the null hypothesis must be rejected. Nevertheless, the investigator may not choose one of the techniques at random. *Different methods of analyzing the data will lead to different conclusions.* In particular, some procedures lead to a finding of no significant difference even though a true difference could be identified by another attack.

Let us illustrate first with some of Kaback's data (29). She administered the group Rorschach to men in certain occupations, and, *inter alia*, compared her groups on the number of popular responses. The mean for accountants is 7.0; for accounting students, 7.3. By the *t*-test, the difference between means is not significant ($P \text{ ca. } .40$). (Point bi-serial r applied to the same data gives the same significance level. Point bi-serial r and t are interchangeable procedures, and there is no merit in testing the hypothesis in both ways.) But if she had chosen the chi-square test, quite proper for her data, Kaback would have found a significant difference between the groups. Chi-square would be applied to compare the proportion of cases in each group having five or more popular responses. From her Table IV, this proportion is 60/75, accountants; 72/75, accounting students. The difference between accountants and accounting students is significant ($P < .01$.) In this and other instances, Kaback disregarded a difference when the null hypothesis could be confidently rejected.

Further illustrative data are taken from Hertz' comparison of Rorschach scores of boys and girls. She tested each possible difference by several statistical devices, yielding results such as those for $M\%$ reproduced in Table II. By any of nine methods, she is informed that the two sex groups differ no more than might two chance samples. By the other computations, she is informed that the difference is significant at the 5% level. If different significance tests disagree, what one concludes depends nearly as much on what procedure one adopts as on the data themselves.

Hertz compared her boys and girls in 46 instances. Each time, she tested the significance of differences between means and between medians. Four times the means differed significantly; five times, the medians differed significantly. But in only one out of 46 comparisons was the difference significant by both methods. It is greatly to Hertz' credit that she saw the applicability of more than one significance test. But conclusions of research will be hopelessly confused and contradictory, unless we can find a basis for choosing between the procedures when one says " 'Tis significant" and the other says " 'Taint."

The choice between comparison of means and medians or between the *t*-test and chi-square cannot be left to the inclination of the experimenter; the whole point of statistical method is to make an analysis freed from subjective judgment. The reason different methods yield different results is that they make different assumptions or try to disclose different aspects of the data. It is therefore important to recognize the ways in which the techniques differ. Differences which are of little

concern in connection with most studies have peculiar importance in Rorschach work. The difficulties which make choice of procedures an important problem arise from three causes: the skewness of Rorschach scores, the complications introduced by ratio scores, and the dependence of Rorschach scores on the total number of responses.

*Choice of Techniques in View of the Inequality of Units
in Rorschach Scales*

Many of the significant Rorschach scores give sharply skewed distributions for most populations. This fact is reported repeatedly (2, 25, 47). Skewness is usually found where many subjects earn 0, 1 or 2 points (i.e., *M*, *FM*, *m*, the shading scores, *CF*, and *C*), and in the location scores *W*, *D*, *Dd*, and *S*. Skewness itself is no bar to conventional significance tests. But in skew distributions the mean and median are not the same. Two distributions may have a significant difference in medians, and not in means (or vice versa) if either is skewed.³ Furthermore, it is doubtful if a satisfactory estimate of $s.d._{mdn}$ can be obtained for a skewed distribution.

Disadvantages of the mean and related procedures. In any statistical computation based on addition of scores (mean, $s.d.$, t , analysis of variance), numerical distances between scores at different parts of the scale are treated as equal. Thus, since the average of 3 *W* and 7 *W* is the same as that of 1 *W* and 9 *W*, these computations assume that a shift 3 *W* to 1 *W* is equivalent to, or counterbalances, a shift 7 *W* to 9 *W*. There is no way of demonstrating equality of units unless one has some knowledge of the true distribution of the trait in question, or a definition of equality in terms of the characteristics of the property being measured. This problem is present in virtually all psychological tools, but other tests yield normal distributions which are assumed to represent the true spread of ability. (On the other hand, *Rorschach interpretation based on clinical experience constantly denies the equality of units for Rorschach scores.*) The average *W* score is near 6, and scores from 1 to 10 are usually considered to be within the normal range. No matter how extremely a person is lacking in *W* tendency, his score cannot go below zero. For one who overemphasizes *W*, the score may go up to 20, 30, or more. A *W* score only six points below the mean may be considered clinically to be as extreme in that direction as a score fifteen points from the mean in the other direction. Munroe (42) has prepared a checklist which shows how units of certain Rorschach scores would have to be grouped in

³ This argument is presented by Richardson (48). In attempting to study differences in medians, Richardson unfortunately uses an incorrect method of determining $s.d._{mdn}$.

order to represent a regularly progressing scale of maladjustment. Her groupings based on clinical experience are of approximately this nature:

W (or $W\%$): 0 (or 1 poor) W response; 1-14%; 15-60%; 61-100%.
 $Dd\%$: 0-9%; 10-24%; 25-49%; 50-100%.
 m : 0-1; 2-3; 4-5; 6 or more.

If these units represent increasing degrees of maladjustment, the raw Rorschach scores do not form a scale of psychologically equal units. It is advisable to accept the clinical judgment on this point, especially in the absence of evidence for the assumption of equal units.

✓ *Use of median and chi-square.* Unlike procedures involving the addition of scores, procedures based on counting of frequencies make no assumption about scale units. In fact, they give the same results no matter how the scale units are stretched or regrouped. The median, or the number of cases falling beyond some critical point (e.g. 10 W), depends only on the order of scores. (This appears to justify the recommendation that counting procedures such as the median be given preference over additive procedures such as the mean in dealing with skew Rorschach distributions.) To test the significance of a difference between two groups, the best procedure is to make a cut at some suitable score, and compare the number of cases in each group falling beyond the cut, using chi-square. This procedure is used by Rapaport (47) and Abel (1). The test of significance of differences between proportions yields the same result (see above). One virtue of cutting scores is that we may test for differences between groups both in the "high" and "low" directions. This is important, since either very high $F\%$ or very low $F\%$, for example, may have diagnostic significance. In the usual analysis based on means, deviations of the two types cancel.

(In contrast to the chi-square method, many tests of significance involve computation of the standard deviation. These include the critical ratio of a difference between means or medians, analysis of variance, and the t -test. In these procedures, great weight is placed on extreme deviations from the mean.) If mean W is 6, a case having 25 W increases Σd^2 (which enters the computation of the $s.d.$) by about 361 points; a case having 15 W increases Σd^2 by about 81 points; and 0 W , by only 36 points. In skewed Rorschach distributions, the few cases with many responses in a category have a preponderant weight in determining σ and the significance of the difference. Whether weighting extreme cases heavily is acceptable depends on whether one considers the difference between 15 W and 25 W to be psychologically large and deserving of more emphasis than, say, the difference from 0 W to 5 W . (Chi-square weights equally all scores below (or above) the cutting point.)

Normalizing distributions. One method used to obtain more equal units is to assume that the trait underlying the score is distributed normally in the population. Raw scores are converted to T -scores which

are normally distributed (35, 67). (This procedure must be distinguished from another conversion, also called a *T*-score, used by Schmidt (54). Scores of the type Schmidt used are not normally distributed.) The effect of normalizing is to stretch the scale of scores as if it were made of rubber. Extreme scores below the median are weighted symmetrically to extreme scores above the median. Thus, in the conversion table prepared by Rieger and used by the writer (10), the median ($6\frac{1}{2} W$) is placed at 100, and a score of 0 *W* is converted to 66, while 28 *W* becomes 134. This in effect compresses the high end of the *W* scale and expands the low end. This conversion does not alter any conclusion or significance test obtained by dichotomizing raw scores and applying chi-square. But the conversion alters markedly any conclusion based on variance or on comparison of means.

There is obviously much merit in using a procedure which leads to a single invariant result, independent of the assumption of the investigator about the equivalence of scores. Even if scores are normalized it is advised that the median be used to indicate central tendency, and chi-square to test significance. (If, for some experimental design, the data must be treated by analysis of variance, the writer believes normalized scores will give results nearer to psychological reality than raw scores, but this judgment is entirely subjective.)

Comparison of mean rank. Attention should be drawn to a new technique invented by Festinger (14) which is peculiarly suitable to the problem under discussion. This method assumes nothing about equality of units or normality of distributions, being based solely on the rank-order of individuals. To test whether two groups differ significantly in a score, one pools the two samples and determines the rank of each man in the combined group. (The mean rank for each group is computed and the significance of the difference is evaluated by Festinger's tables. The method has not yet been employed in Rorschach research.)

The Festinger method and chi-square are not interchangeable. Which should be used depends on the logic of a particular study. Chi-square answers such a question as "Does Group A contain more deviates than Group B in the score being studied?" The Festinger method gives weight to differences all along the scale, and therefore asks whether the two groups differ, all scores being considered. In one study, absence of *M* is quite important but differences in the middle of the range have no practical importance. In another study, differences all along the scale are worth equal attention.

The Festinger method appears to have the advantage of greater stability for small samples. Chi-square is much easier to use in samples of 30 or more per group. The Festinger method is not useful when there are numerous ties in score. Further experience with the new method may disclose other important distinctions.

Significance Tests Compared with Estimates of Relationship

Some investigators have perhaps not conveyed the full meaning of their findings to the reader because of a failure to distinguish between tests of the null hypothesis, and estimates of the probable degree of relationship between two variables. The former type of result is a function of the number of cases, whereas the latter is not, save that it becomes more trustworthy as more cases are included. When an investigator applies chi-square, the *t*-test, or the like, he determines whether his observations force him to conclude that there is a relationship between the variables compared. But if the degree of relationship is moderately low, and the number of cases small, the null hypothesis is customarily accepted even though a true relationship exists. It is proper scientific procedure to be cautious, to reject the hypothesis of relationship when the null hypothesis is adequate to account for the data. (But in Rorschach studies, where sample size has often been extremely restricted, nonsignificant findings may have been reported in a way which discouraged investigators from pursuing the matter with more cases.)

The study of McCandless (36) is a case in point. McCandless compared Rorschach scores with achievement in officer candidate school. In each instance save one, the *t*-test showed *P* greater than .05 that the difference would arise in chance sampling. But the samples compared contained only thirteen men per group. Under these circumstances, it would take a sharply discriminating score to yield a significant difference. If the sample size was raised to about 50 per group, and the differences between groups remained the same, twelve more of McCandless' thirty significance tests would be significant at the five percent, or even the one percent, level. When more cases are added, the differences will certainly change and most of them will be reduced in size. In fact, the writer believes, on the basis of other experience with statistical comparisons of the Rorschach with grades, that McCandless' negative findings are probably close to the results which would be found with a larger sample. But the point is that McCandless, and other investigators using small *N*'s, have submitted the Rorschach to an extremely, perhaps unfairly, rigorous test. One way to compensate for the necessary rigor of proper significance tests is to also report the degree of relationship. A chi-square test may be supplemented by a contingency coefficient or a tetrachoric *r*. A *t*-test may be supplemented by a bi-serial *r*, or point bi-serial (not to determine significance, as Kaback used it, but to express the magnitude of the relationship). Sometimes reporting the means of the groups and their standard deviations, to indicate the degree of overlapping, is an adequate way to demonstrate whether the relationship looks promising enough to warrant further investigation.

To restate the problem: the investigator always implies two things in a comparison of groups: (a) that he considers the null hypothesis is definitely disproven by his data, or else that the null hypothesis is one way to account for the data, and (b) in case the null hypothesis still

remains tenable, that he does or does not judge further investigation of the question to be warranted. He can never prove that there is no relationship. So, if his data report a non-significant difference, he must judge whether the difference is "promising" enough to warrant further studies. This judgment is not reducible to rules in the way the significance test is. Whether to recommend further work depends on the difficulty of the study, on the probable usefulness of the results if a low order of relationship were definitely established by further work, and in the investigator's general confidence that the postulated relationship is likely to be found.

Methods of Partialling Out Differences in R

The usual approach when comparing groups is to test the differences in one score after another, and then to generalize that the groups differ in the traits to which the scores allegedly correspond. The various scores, however, are not experimentally independent—a man's total record is obtained at once, and his productivity influences all his scores. If two groups differ in *R*, they may also differ in the same direction in *W* (whole responses), *D* (usual details), and *Dd* (unusual details).

Thus consider the Air Force data in Table III.

TABLE III

RORSCHACH SCORES COMPARED TO SUCCESS IN PILOT TRAINING (21, p. 632)

<i>Rorschach Score</i>	<i>Mean of Successful Cadets</i>	<i>Mean of Unsuccessful Cadets</i>	<i>Bi-serial r</i>
<i>R</i>	18.5	15.8	.14
<i>W</i>	9.2	7.3	.24
<i>D</i>	7.1	6.7	.03
<i>W%</i>	60.2	55.8	.08
<i>D%</i>	31.7	37.6	-.15

The first group has more responses than the second. From the means in *W* and *D*, it would appear that the first group has more *W* tendency than the second, but is equal in *D*. But when responsiveness is controlled by converting scores to percentages, the difference in *W* becomes small and the second group is shown to be stronger than the first in emphasis on *D*.

The most striking illustration of this difficulty is Goldfarb's comparison of obsessionals and normals. The obsessional group averages 55 *R*; the normals, 14. Under the circumstances, it is not at all informative to proceed to test *W*, *D*, and *Dd*; all differ significantly in the

*bal
sham*

same direction. One learns nothing about differences between groups in mental approach, which is the purpose of considering these three scores. Most of Goldfarb's other comparisons also merely duplicate the information given by the test in *R*, that is, that the obsessionals are more productive. Although the discrepancy between the groups in *R* is unusually striking in Goldfarb's group, it is present to a lesser but significant degree in a great number of other studies, including those of Buhler and Lefever (5), Hertzman (26), Kaback (29), Margulies (38), and Schmidt (54).

(A similar problem complicated Beck's comparison of schizophrenics and normals on *D*. The means were 19.0 and 19.9, respectively; the σ 's were 13.5 and 9.9. Beck comments as follows:

The small difference is accentuated in the very small Diff./S.D. diff: 0.34. There is, however, probably a spurious factor in this small difference. The ogives give us a hint: up to the eighty-second percentile, the curves run parallel, with that for controls where we should expect it, higher. Above this point, the schizophrenics' curve crosses over, and continues higher, and more scattering, as we should expect from the S. D. The spurious element lies undoubtedly in the fact that the schizophrenics' higher response total would necessarily increase the absolute quantity of *D*, since these form the largest proportion of responses in practically all records. Absolute quantity of details is then no indicator of the kind of personality we are dealing with. . . . The medians for *D* are 14.46, 17.2 (2, pp. 31-32).

When one makes several significance tests in which the difference in *R* reappears in various guises, one becomes involved in a maze of seemingly contradictory findings. And interpretation tempts one to violate the rule of parsimony, that an observed difference shall be interpreted by the fewest and simplest adequate hypotheses. To answer the question, how do obsessionals and normals differ? [it is simpler to speak of the former as more productive than to discuss three hypotheses, one for each approach factor. And one may certainly criticize Hertzman and Margulies (27) for interpreting differences in *D* and *Dd* between older and younger children as showing the former's greater "cognizance of the ordinary aspects of reality" and greater concern with facts.] The older group gives twice as many *R*'s as the former, which is sufficient to account for the remaining differences.

One might argue that *R* is resultant rather than cause, and that the differences in *W*, *D*, *Dd*, etc. are basic. But the Air Force demonstration that *R* varies significantly from examiner to examiner (21) suggests strongly that responsiveness is a partly superficial factor which should be controlled.

Only two studies examine their data explicitly to determine if differences in

other categories could be explained in terms of responsiveness alone. Werner (68) found a significant difference in $dd\%$ between brain-injured and endogenous defectives. But the latter gave significantly more R 's. He therefore counted only the first three responses in each card, and arrived at new totals. With R thus held about constant, he found the dd difference still marked and could validly interpret his result as showing a difference in approach.

Freeman and others (17) found that groups who differed in glucose tolerance also differed significantly in R . After testing differences in M and $\text{sum } C$ on the total sample, they discarded cases until the two subsamples were equated in R . Since differences between the groups in M and C were in the same direction even when R was held constant, they were able to conclude with greater confidence that glucose tolerance is related to M and C .

After differences in R are tested for significance, it is appropriate to ask what other hypotheses are required to account for differences in the groups. But these other hypotheses should be independent of R ; otherwise one merely repeats the former significance test and obscures the issue. The usual control method is to divide scores by R , testing differences in $W\%$, $D\%$, $M\%$, $A\%$, $P\%$, etc. Such ratios present serious statistical difficulties discussed in the next section. Moreover, these formulas fail to satisfy the demand for independence from R . There may be correlation between R and $W\%$, etc. (For a sample of 268 superior adults from a study by Audrey Rieger, the writer calculates these r 's: $W\% \times R$, $-.45$, $M\% \times R$, $.03$, $F\% \times R$, $.06$. In the latter two cases, there is no functional relation of the percentage with R , but the distributions are heteroskedastic. $\sigma_{W\%} = 3.30$ when R 5-19 (74 cases) but 2.09 when R 40-109 (82 cases). The corresponding sigmas for $M\%$ are 3.85 and 3.35; for $F\%$, 3.23 and 2.29. Only $M\%$ is really independent of R .)

One may control differences in R by other methods, provided many cases are available. One procedure is to divide the samples into subgroups within which R is nearly uniform (e.g., R 20-29), and make significance tests for each such set. A method which requires somewhat fewer cases is to plot the variable against R for the total sample or a standard sample, and draw a line fitting the medians of the columns. This may be done freehand with no serious error. Then the proportion of the cases in each group falling above the line of medians may be compared by chi-square.

Difficulties in Treating Ratios and Differences

More than any previous test in widespread use, the Rorschach test has employed "scores" which are arithmetic combinations of directly counted scores. One type is the ratio score, or the percentage in which the divisor is a variable score. Examples are $W:M$, $M:\text{sum } C$, W/R

($W\%$), and F/R ($F\%$). The other type of composite is the difference score, such as $FC - (CF + C)$. In clinical practice, scores of this type are used to draw attention to significant combinations of the original scores; the experienced interpreter thinks of several scores such as FC , CF , and C , at once, placing little weight on the computed ratio or difference. When these scores are used statistically, however, there is no room for the flexible operation of intelligence; the ratios are treated as precise quantities.

It may be noted in passing that a few workers (e.g., 63) appear to assume that $\text{Mean } a / \text{Mean } b$ is the same as $\text{Mean } \frac{a}{b}$. This is of course not true; the mean of the ratios and the ratio of the means may be quite unequal. One cannot, as Kaback did (29, pp. 33, 53, 55), assume that if the ratio of the means is greater for one group than another, the groups differ in the ratio scores themselves. The reader may convince himself by computing the mean ratio for each of the following sets of data in which $\text{Mean } a / \text{Mean } b$ is constant:

$$\frac{0}{2}, \frac{2}{4}, \frac{4}{6}, \frac{6}{8}, \frac{8}{10}; \frac{0}{6}, \frac{2}{8}, \frac{4}{2}, \frac{6}{10}, \frac{8}{4}; \frac{0}{10}, \frac{2}{8}, \frac{4}{6}, \frac{6}{4}, \frac{8}{2}.$$

One difficulty with ratio scores is their unreliability. Consider a case with 5 W , 1 M . The ratio $W:M$ is 5. But M is a fallible score. On a parallel test it might shift to 0 or to 2. If so, the ratio could drop to $2\frac{1}{2}$, or zoom to infinity; such a score is too unstable to deserve precise treatment. The unreliability of another ratio is illustrated in Thornton and Guilford's data (61). The reliabilities were, in one sample, .92 for M , .94 for C , but .81 for M/C . In a second sample, the values were .77, .65, and .31. If unreliable ratios are added, squared, and so on, one commits no logical error, but psychologically significant differences become overshadowed by errors of measurement.

Ratios based on small denominators are in general unreliable (7). $W\%$ is unreliable for a subject whose R is 12, but relatively reliable for a case whose R is 30. In the former case, addition of one W response raises $W\%$ by 8; in the latter, by 3%. Errors of measurement always reduce the significance of differences by increasing the within-groups variance. A significant difference in $W\%$ might be found for cases where $R > 25$. A difference of the same size might not be significant for cases where $R < 25$ because of the unreliability of the ratio. If the significance test were based on all cases combined, the difference might be obscured by the unreliability of the ratios in the latter group. One possible procedure is to drop from the computations all cases where the denominator

is low. (If there is a significant difference even including the unreliable scores, this need not be done.)

The issue of skewness must again be raised. In the $M: \text{sum } C$ ratio, all cases with excess C fall between zero and 1. Those with excess M range from 1 to ∞ . The latter cases swing the mean and sigma. Following the argument of a preceding section, it is injudicious to employ statistics based on the mean and standard deviation, as McCandless (36) did. By such procedures, different conclusions would often be reached if both $M: \text{sum } C$ and $\text{sum } C: M$ were tested. Procedures leading to a chi-square test are to be recommended, as illustrated in several studies (Rapaport, 47, pp. 251; Rickers, 49; etc.) Another solution, less generally suitable, is to convert ratio scores to logarithmic form to obtain a symmetrical distribution (61).

A hidden assumption in ratios and differences is that patterns of scores yielding equal ratios (or differences) are psychologically equal. Thus, in $W\%$ the same ratio is yielded by 2 W out of 10 R , 8 W out of 40, and 20 W in 100 R . One can always define and manipulate any arbitrary pattern of scores without justifying it psychologically, but better conclusions are reached if the assumption of equivalence is defensible. (The regression of W on R is definitely curved. A person with 2 W out of 10 R is low in W tendency, since it is very easy to find two wholes in the cards.) Only people with strong tendency and ability to perceive wholes can find 20 W in the ten cards, regardless of R . As R rises above 40, W seems to rise very little; the additional responses come principally from D and Dd . The resulting decline in $W\%$ reflects a drive to quantity, rather than a decreased interest in W (cf. 47, p. 156). Put another way: a strong drive to W can easily lead to 90 or 100% W when $R < 15$; but such a ratio in a very productive person is unheard of. If the regression of a on b is linear and a close approximation to $(a/b) = \text{some constant}$, ratios may be used as a score with little hesitancy. Otherwise the ratio is a function of the denominator.

This factor is recognized by Munroe, who indicates repeatedly in her checklist that the significance of a particular ratio depends on R . Thus 30-40% M is rated + if $R = 10$, but 16-29% is rated + if $R = 50$. Numerically equal Rorschach ratios, then, are not psychologically equal. Rapaport reflects the same point in testing differences between groups in W/D . Instead of applying chi-square to the proportions having the ratio 1:2 or lower, he adjusted his standard.

In records where R is too low or too high, we took cognizance of the fact that it is difficult not to get a few W 's and difficult to get too many. Thus, in low

whol
dist.
w/W's

R records the 1:2 norm shifted to a "nearly 1:1" while in high *R* records, the 1:2 norm shifted to a 1:3 ratio (47, p. 134).

This adjustment was evidently done on a somewhat subjective basis, and is therefore not the best procedure. It is unfortunate that most other workers have unquestionably assumed that a given score in *W*%, *M*%, or $FC - (CF + C)$ has the same meaning regardless of *R*.

At best, ratio- and difference-scores introduce difficulties due to unreliability and to assumptions of equivalence. There is a fairly adequate alternative which avoids statistical manipulation of ratios entirely. One need only list all significant patterns, and determine the frequency of cases having a given pattern. Thus *M: sum C* can be treated in these categories: coartated (*M* and *C* 2 or below); ambiequal, *M* or *C* < 2, *M* and *C* differ by 2 or less; introversive, *M* exceeds *C* by 3 or more; extra-tensive, *C* exceeds *M* by 3 or more. Any other psychologically reasonable division of cases may be made, and significance of differences tested by chi-square, provided that the hypothesis is not chosen to take advantage of fluctuations in a particular sample. Even this method, however, does not escape the criticism that a given pattern of two scores, such as 3 *M*, 3 *C*, has different significance in records where *R* differs greatly. To cope with this limitation, the pattern tabulation procedure is suggested later.

A detailed consideration of certain work by Margulies is now appropriate, since it affords an illustration of many problems presented above. Her study of the *W:M* ratio employed a procedure almost like that just recommended, but with departures which are unsound. Margulies compared Rorschach records of adolescents having good and poor school records (38). Only her 21 successful boys and her 32 unsuccessful boys need be considered here. She was interested in comparing them on the *W:M* pattern, in view of Klopfer's belief that this ratio indicates efficient or inefficient use of capacity. She not only tested her data in several ways, but reported the data so that other calculations can be made. Table IV reproduces a part of her data, and shows the results of seven different procedures for determining the significance of the difference.

It should be noted first that Yates' correction is essential for tables with 1 d.f. and low frequencies; in each case where it is applicable, the correction lowers the significance value importantly. Second, attention may be turned to the use of chi-square to test differences between two distributions. Even if more cases were available, it would be unwise to apply chi-square to the distribution cell-by-cell (Procedures 2, 3), since this procedure ignores the regular trend from class-interval to class-interval. Instead, the distribution should be dichotomized. Therefore, procedure 5 is preferable to 2, and 6 is preferable to 3. It will be noted that these recommended procedures indicate higher significance than the tests in which the distributions are compared cell-by-cell.

Margulies is one of the few writers to note the unsoundness of assuming that equal ratios are equal. She pointed out that 20 *W*: 10 *M* is not psychologically

TABLE IV

RESULTS OBTAINED WHEN A SET OF DATA IS TREATED BY A VARIETY OF PROCEDURES
(Data from Margulies, 38, pp. 23, 26, 44)

Distribution I			Distribution II			Distribution III		
Number of M	Successful boys	Unsuccessful boys	W/M ratio	Successful boys	Unsuccessful boys	Pattern of W and M	Successful boys	Unsuccessful boys
3 or more	5	5	<1	1	1	W < 6, M 0-1	0	10
2	9	8	1.00	0	2	W < 6, M > 1	8	2
1	3	11	1.1-2.9	8	5	W > 5, M 0-1	7	9
0	4	8	3.0-4.9	5	7	W 6-10, M 2	3	7
			>4.9	3	9	W 6-10, M > 2	1	4
			∞ (W/0)	4	8	W > 10, M > 1	2	0

similar to 2 W: 1 M, and she demonstrated that the regression of M on W is significantly curvilinear. She therefore was properly critical of procedures such as 3 and 6. She next turned to the scatter diagram of M and W, and found successful boys predominating in some regions, and unsuccessful boys in others. After grouping scores into regions as shown in Distribution III, she divided the surface into two areas, one area including cases where W is 0 to 5 and M is 2 or over, plus cases where W is 6 to 10 and M is 3 or over, plus cases where W is over 10 and M is 2 or over. In other words, instead of testing whether the groups are differentiated by a cut along the straight line $M = 2$ (Procedure 5), she made her cutting line an irregular one. This hypothesis, tested in Procedure 7, gave apparently quite significant results. The results are of little value, however, since the hypothesis was "cooked up" to fit the irregularities of these specific data. In the cells where W is 6 to 10, and M is 2, there happens to be a concentration of unsuccessful boys. But to draw the cutting line irregularly to sweep in all areas where the unsuccessful predominate is a type of gerrymandering which vitiates a significance test. Hundreds of such irregular lines might be drawn. Therefore, it would be expected that in any sample some line could be found yielding a difference "significant" at the 1% level. At best, the irregular line sets up a hypothesis which, if found to yield a significant difference in a new and independent sample, could be taken as possibly true.

The law of parsimony enters this problem. Wherever a set of data may be explained equally well by two hypotheses, it is sound practice to accept the simpler hypothesis. Irregular cutting lines, and explanations in terms of patterns of scores, are sometimes justified and neces-

TABLE IV—(continued)

Type of analysis	Procedure	Result	P	Results with Yates' correction	
				χ^2	P
Central tendency	1. Significance of difference in mean M	CR = .70*	.48
Cell-by-cell comparison	2. Chi-square applied to Distribution I (3 d.f.)	$\chi^2 = 3.78^{***}$	ca. .30
	3. Chi-square applied to Distribution II (5 d.f.)	$\chi^2 = 5.30^*$	ca. .40
	4. Chi-square applied to Distribution III (5 d.f.)	$\chi^2 = 17.73^*$	<.01
Dichotomy	5. Chi-square applied to number of cases with $M > 1$ (Dist. I)	$\chi^2 = 3.46^{**}$.06	2.54**	.11
	6. Chi-square applied to number of cases with $W/M > 3$ (Dist. II)	$\chi^2 = 1.86^{**}$.18	1.13**	.30
Frequency of selected patterns	7. Chi-square applied to frequency having $M > 1$ if $W > 6$ or > 10 ; having $M > 2$ if $6 < W < 10$ (Dist. III)	$\chi^2 = 6.58^{**}$.01	5.13**	.03

* Computed by Margulies.

** Computed by the writer.

*** Computed by the writer. Margulies reports 3.64.

sary. But in this case the difference between the groups is explained as well by the hypothesis that the successful boys give more M 's as by any non-spurious test of the W ; M relationship. Therefore, procedure 5 is the soundest expression of the significance of the Margulies data. With more cases, this difference might be found to be truly significant.

In the above analysis, we find again that different procedures, more than one of which is mathematically sound, give different conclusions. The results from chi-square are less compatible with the null hypothesis than is the critical ratio. Chi-square applied to a dichotomy gives evi-

dence of a possible relationship whereas chi-square applied to the frequency distribution does not. Attention is again drawn to the necessity of regarding with great suspicion any significance test based on a complex hypothesis set up to take advantage of the fluctuations of frequencies in a particular sample. Finally, it is noted that explanations in terms of ratios and patterns should not be sought unless they can account for observed differences more completely than can hypotheses in terms of single scores.

TREATING PATTERNS OF SCORES

Rorschach workers continually stress the importance of considering any score in relation to the unique pattern of scores for the individual. While this is done in clinical practice, there is no practical statistical procedure for studying the infinite complex interrelations of scores and indications on which the clinician relies. Instead of considering the individual patterns, the statistician can at best study certain specific patterns likely to occur in many records. A pattern can be exceedingly complex; there is no statistical reason to prevent one from studying whether (for example) more men than women show high-*S*-on-colored-cards-accompanied-by-emphasis-on-*M*-and-excess-of-*CF*-over-*C*. The only limitation the statistical approach imposes is that the same pattern of scores must be studied in all cases.

Patterns of scores may be considered by means of composite scores, by definition of significant "signs," and by the pattern-tabulation method. The composite score is simply an attempt to express, in a formula, some psychologically important relationship. Examples include the *M*: *sum C* ratio, and the more complex composites developed by Hertz or Rapaport. (These scores may be treated statistically like any score on a single category, although most of them are ratios or differences and suffer from the limitations already discussed.)

Comparing incidence of "signs." The "signs" approach has been widely used. It is simple and well-adapted to the Rorschach test. Normally, an investigator identifies some characteristic of a special group, such as neurotics, from clinical observation. Then this characteristic is defined in a sign, i.e., a rule for separating those having the characteristic. One such sign, for example, is $FM > M$. After the investigator hypothesizes that some sign is discriminative, the necessity arises for making a test of significance to see if the sign is found more often in the type of person in question. (One may soundly compare a new sample of the diagnosed group with a control sample by noting the frequency of the sign in each group and applying chi-square. This procedure is illustrated in studies by Hertzman and Margulies (38), and Ross (50).)

The investigator may invent his own signs, if he follows due precautions to avoid misleading inflation of probabilities. Often it is easier and equally wise to use a predetermined set of signs. The most useful set of signs available at present is the Munroe checklist. She has identified numerous ratios and patterns of scores which she considers significant of disturbance in her subjects (adolescent girls). (She has stated that she does not think of her method as a set of signs (41), but the difference between her list and others appears to be (a) that it provides an inclusive survey of all deviations in a record and (b) that the list is designed as a whole to minimize duplication from sign to sign. There is no reason why two groups may not be compared by applying the checklist to every record, and then comparing the groups on the frequency with which they receive each of the possible checks.) Chi-square is the proper significance test, as used in one of Munroe's studies (43). The Munroe signs sometimes are simply defined (e.g. $P-$ is 0 or 1 popular response), but some involve patterns of several scores (thus the sign $FM+$ is defined in terms of FM , M , and R).)

Pattern tabulation. Pattern tabulation is a method devised by Cronbach for the study of relations between two or three scores (10). It has the advantage of permitting one to study the distribution of patterns in a group. To deal with any set of three scores, e.g. W , D , Dd , one normalizes the three scores for each person, and considers the resulting profile. The profile is expressed numerically in terms of the deviation of the converted scores from their average for each person. These three scores can be plotted on a plane surface, and the resulting scattergram shows the distribution of patterns in a group. If two groups are compared, any type of pattern found more commonly in one group than another can be identified, and the difference in frequency tested by chi-square. The significance level for rejecting the null hypothesis must be set conservatively, as this method involves many implied significance tests. An analysis of variance solution is also possible but not recommended in view of the fact that distributions of patterns are often non-normal.

This method cannot consider hypotheses involving more than three scores at once. It functions best when the three scores are equally reliable and equally intercorrelated. It encounters difficulty due to the fact that some Rorschach scores are unreliable, since any serious error of measurement in one score throws an error into the profile. The method does, however, appear flexible and especially useful for such meaningful patterns as $W-D-Dd$ and M -sum $C-F$.

Another group of procedures leading to composite formulas for discriminating groups is treated in the next section.

DISCRIMINATION BY COMPOSITE SCORES

In many problems, it is desired to use the Rorschach to discriminate

between two groups. Thus one might seek a scoring formula to predict pilot success, or a "neurotic index" to screen neurotics from a general population. The methods used to arrive at composite scores are the checklist, the multiple regression equation, and the discriminant function.

Checklist scores. The checklist consists of a set of signs. Each person is scored on the checklist and the total number of signs or checks is taken as a composite score. This method has had considerable success, notably in Munroe's study (42) and in the formula of Harrower-Erickson and Miale for identifying insecure persons. There are no serious statistical problems in the use of checklists. The total score can be correlated (though eta may be preferable to r). Differences between groups may be tested for significance, preferably by chi-square. Chi-square is advised because a difference in the non-deviate range is rarely psychologically significant; the investigator is usually concerned with the proportion of any group in the deviate range. Buhler and Lefever justifiably apply analysis of variance to their checklist score, to study its ability to differentiate clinical groups (5).

Problems do arise, however, in developing checklist scores. A common method is to compare two groups on one raw score after another, noting where their means differ. Each score where a difference arises is then listed as a sign, and counted positively or negatively in obtaining the checklist score for each case. This method takes advantage of whatever differences between samples arise just from accidents of sampling. If sample A exceeds B in mean M , allowing one point in the total score for high M will help discriminate A's and B's. In this sample, the A's will tend to earn higher checklist scores. But often in a new sample such a difference will not be confirmed, and the M entry in the composite will not discriminate.

One study employing the sign approach should be pointed out to Rorschach workers. Davidson (12) sought to determine the relationship between economic background and Rorschach performance in a group of highly intelligent children. Her treatment of data is noteworthy because of the flexibility of her procedures; statistics are applied with great intelligence, new procedures being adopted for each new type of comparison. While the reviewer disagrees with some of the judgments she made in selecting procedures, her treatment is free from overt errors and well worth study by other Rorschach investigators.

Davidson divided her 102 cases among seven economic levels. She studied the Rorschach performance in various ways. First, she made a clinical analysis of each child, and placed him in one of nine categories (introvert adjusted, childish, constricted, disturbed, etc.). The distribution which resulted is a 7×9 table. Recognizing that the expected frequency in each cell is quite small, she combined groups to form a 3×3 table before applying the chi-square test for significance. This same type of condensation would have been advisable in some other comparisons she made, such as that between personality pattern and IQ.

Davidson next applied a list of signs, and obtained for each case the total number of signs of maladjustment. The number of signs was correlated with economic level, and the correlation was shown not to differ significantly from zero. She tested the significance of the difference in mean number of signs by the critical ratio. These procedures appear well suited to her data. A third attack on the data treats one Rorschach score at a time. Here Davidson placed her cases in seven categories, ranging from highest to lowest economic level. By analysis of variance, she demonstrated that differences among the seven groups were significant only for a few of the scores. The application of analysis of variance to continuous data appears to have been an unwise decision. Analysis of variance, like chi-square or eta applied to a variable divided in several categories, ignores the order of the categories. Consider the following set of means in the score $M - \text{sum } C$:

Economic level	1	2	3	4	5	6	7	Total
Mean score	1.17	1.86	1.29	0.96	-0.75	-0.13	-0.71	0.63

The downward trend from Group 1 to Group 7 gives great support to the hypothesis that this score is related to economic level. Analysis of variance estimates significance without considering this trend; the same significance estimate would be arrived at if Group 2 had had the mean of -0.13 and Group 6 the mean of 1.86 . Davidson might have computed the correlation between each score and the economic level, but the skewness of some Rorschach scores weighs against this suggestion. The simplest procedure for testing this trend is to split the group into a 2×2 table by combining adjoining categories in the economic scale, and dichotomizing the Rorschach score at a convenient point. Chi-square would then give the significance estimate. Such a procedure might have yielded significant differences in several instances where Davidson found none.

In justice to Davidson, it should be repeated that her data have been singled out for critical comment because of their exactness and completeness, rather than because they were improperly handled. The foregoing suggestions point to ways in which she might have arrived at additional important findings.

The multiple regression formula. A limitation of check lists is that they are simple additive combinations of signs which individually discriminate. But in such a composite a given trait may enter several times if it is reflected in several signs, and thus have greater proportionate weight than it deserves. The checklist method does not allow for the possibility that certain signs may reinforce each other to indicate more severe maladjustment than is indicated by a combination of two other non-reinforcing signs, or for the possibility that two signs which are individually unfavorable may operate to neutralize each other. Multiple regression and the discriminant function are more powerful procedures than the usual checklist score, because they consider the intercorrelations of scores and weight them accordingly.

By multiple correlation, one arrives at a regression equation which assigns weights to those variables which are correlated with a criterion and relatively uncorrelated with each other. This formula may be used

to predict or to discriminate between groups. One such formula is that of the Air Force, used in its attempt to predict pilot success (21):

$$2(Dd+S\%)+6 FM+8 W-1.5 D\%+R-(VIII-X\%).$$

Multiple correlation does not seem especially promising for Rorschach studies. Even such an elaborate formula as that above turns out to have little or no predictive value when applied to a fresh sample. (Even if it were stable, any formula of this type must assume that strength in one component compensates linearly for weakness in another.) In this formula, emphasis on *Dd* would cancel weakness in *FM*, in estimating a man's pilot aptitude. It is most unlikely that the factors cancel each other in the personality itself. The simple linear regression formula provides an efficient weighting if the assumption of linear compensation is valid, but interrelations between aspects of personality are probably far too complex to be adequately represented in this way. The most that can be said for a regression formula is that, when derived on large samples (and this may require 5000 cases), it is a more precise prediction formula than the simple checklist score can be. It cannot hope to yield very accurate predictions if interrelations within personality are as complex as Rorschach interpreters claim.

The discriminant function is a relatively new technique giving a formula which will separate two categories of men as thoroughly as possible from a mixed sample. It would be used to develop an effective index for separating good from poor pilots (not for predicting which man will be best, as the regression formula does) or for distinguishing organics and feeble-minded. A practical procedure for dealing with multiple scores has just been published by Penrose (45), and has not been employed in Rorschach research. It appears likely to have real value in studies comparing different types of subjects.

Like the regression formula, however, the discriminant function provides a set formula. In this formula, it is assumed that one factor compensates for or reinforces weakness in another factor. The interactions within personality are probably too complex to be fully expressed by linear or quadratic discriminant functions.

CORRELATION AND RELIABILITY

Correlations of scores. For one purpose or another several studies have tried to show the relationship between the several Rorschach scores or between Rorschach scores and external variables. The conventional procedure for showing that two characteristics are associated is to compute a product-moment correlation between the variables. This has been done by Kaback (29), Vaughn and Krug (64), and others.

(This method is unable to show the full relationship between variables when the regression of one on the other is curvilinear. Such a regression

often occurs when one variable or both have a sharply skewed distribution. In fact, Vaughn and Krug note that one of their plots is curvilinear. The extent to which association may be underestimated is suggested by the following data. The data used are taken from tests administered individually by Audrey Rieger to several hundred applicants for employment, usually for managerial or technical positions. The tests were carefully scored by the Beck method. Generalization from the data must be limited because the group is not a sample of any clearly defined population. For 268 men, the product-moment correlation between D and Dd is .735. The curvilinear correlations are η_{DdD} , .785; η_{DDd} , .823. There is significant curvilinearity. If D and Dd are normalized, the regression becomes linear except for the effect of tied scores where $Dd=0$; for the converted scores, $r=.767$.

Brower employs rank-difference correlations in comparing certain Rorschach scores to physiological measures (3). This is a useful method for small samples and is equally sound for linear and non-linear regressions. Thus, a rank-correlation of W/M with another score is the same except for sign as the correlation for the inverted ratio M/W , but the product-moment correlations are far different.

The rank method does have the disadvantage of weighting heavily the small and unreliable differences in the shorter end of skew distributions, where many cases have the same rank. This might lower the correlations for a score like Fc , but is not a difficulty with scores distributed more symmetrically over a wide range, such as F or $VIII-X\%$. Normalizing has the same disadvantage. This is a reflection of the inability of the test to discriminate finely among cases in the modal end of a severely skewed distribution.

Reliability coefficients. Test reliability is ordinarily estimated by the retest or the split-half method. These methods are not very appropriate for the Rorschach test, the former because of memory from trial to trial, the latter because the test cannot be split into similar halves. Nevertheless, both methods have been used in the absence of better procedures.

The split-half method introduces a statistical problem which not all investigators have noted, namely, that the Spearman-Brown formula must not be applied to ratios with variable denominators such as $W\%$ and $M/\text{sum } C$. Methods for estimating the reliability of ratio scores have been treated elsewhere (7, 8), but these procedures are not useful when the denominator is relatively unreliable (as in $M/\text{sum } C$).

It is desirable to estimate reliability of scores separately for records of varying length. Vernon (65) found that Rorschach scores were much more reliable for cases where $R>30$ than when $R<30$. This implies that it is unsatisfactory to estimate just one reliability coefficient for a

group with varied R . Instead, the standard error of measurement of W , or $W\%$, should be determined separately for cases where $R=10-15$, $R=15-25$, $R=25-35$, or some such grouping.

The reliability of patterns of scores is a difficult problem. If both M and W were perfectly reliable, any pattern or combination based on the two scores would also be perfectly reliable. But these scores are unstable; subjects vary from trial to trial in M or W or both. Nevertheless, Rorschach users insist that the "pattern" of scores is stable. If there is any substance to this claim, it means that certain definable configurations of the scores are stable even though the separate scores are not. The configurations may be as simple as the W/M ratio or may be complex structures of several scores. One may establish the reliability of any composite score by obtaining two separate estimates from independent trials of the test.

The method of determining reliability by independent estimates has rarely been used. A study by Kelley, Margulies, and Barrera (30) is of interest, even though based on only twelve cases. The Rorschach was given twice, and between the trials a single electroshock was given, reportedly sufficient to wipe out memory of the first trial without altering the personality. In the records so obtained, R shifted as much as 50 per cent from trial to trial, and absolute values of some other scores shifted also. In several cases where scores shifted, it can be argued that the *relationship* between the scores did not shift and that the two records would lead to similar diagnoses. The authors made no attempt at statistical treatment. Probably this ingenious procedure will rarely be repeated. Useful studies could certainly be made, however, by comparing performance on two sets of inkblots without shock (cf. Swift, 57). Even if the two sets are not strictly equivalent, the data would indicate more about the stability of performance than any methods so far employed.

At first glance, it appears logical to set up composite scores, obtain two separate estimates, and correlate them. Even this is unsuitable for Rorschach problems, however. As pointed out before, a given ratio such as 20% W or W/M 2.0 has different meaning in different records, depending on the absolute value of W . The pattern might conceivably be defined by a curvilinear equation, but this becomes unmanageable, especially as several variables enter a single pattern. The problem is one of defining when two patterns are psychologically similar, and of defining the magnitude of the difference when they are not equivalent. No one would contend that the W/M balance is unchanged if a subject shifts from 12 W : 2 M to 60 W : 10 M . The problem is to define and measure the balance in a numerical way. The approach pattern $W-D-Dd$ has three dimensions. If we wish to estimate reliability by comparing two sets of these three scores we have a six-dimensional array, for which no present methods are adequate. So far, even the

pattern-tabulation method reduces such data only to four dimensions, which leaves the problem still unmanageable. All that can be recommended is that additional attention be given to this challenging problem. We can now obtain adequate evidence on the stability of Rorschach patterns only by such a method as Troup's (62), discussed in the first section of this paper. It will be recalled that she had two sets of records interpreted clinically, and employed blind-matching to show that the inferences from the Rorschach remained stable.

Two unique but entirely unsound studies by Fosberg (15, 16) employed a novel procedure to estimate the reliability of the total pattern. He gave the test four times, under varied directions. He then compared the four records for each person. In one study he used chi-square to show that the psychograms for each person corresponded. But this statistical test merely showed that the D score in record 1 is nearer to D in record 2 than it is to W , C , or other scores. That is, he showed that the scores were not paired at random. But, since each score has a relatively limited range for all people—i.e., D tends to be large, m tends to be small, etc.—he would have also obtained a significantly large chi-square if he had applied the same procedure to four records from *different* persons. One may also point out that finding a P of .90 does not prove that two records do come from the same person, but only that the null hypothesis is tenable, or possibly true. Fosberg's second study, using correlation technique, is no sounder than the first. Here the two sets of scores for one person were correlated. That is, pairs of values such as $W_1 - W_2$, $D_1 - D_2$, etc. were entered in the same correlation chart. As before, the generally greater magnitude of D causes the two sets to correlate, but high correlations would have been obtained if the scores correlated came from two different subjects.

Objection must also be made to several procedures and inferences of Buhler and Lefever (5), in their attempts to demonstrate the dependability of their proposed Basic Rorschach Score. (1) They used the split-half method on the total score, by placing half the signs in one list, the other half in a second list, and scoring each person on both lists (5, p. 112). They then correlated the two halves to indicate reliability. Because the correlation was computed on cases used to determine the scoring weights for the items, the resulting correlation is spuriously high. Even if new cases were obtained, the split-half method would be incorrect because the checklist items are not experimentally independent. A single type of performance enters into a great number of separately scored signs (in their checklist, M affects items 1, 2, 5, 6, 7, 8, 10, 11, 12, 51, 52, 53, 86, 93, 94, 95, 96, 99, 100, 101, and 102). A "chance" variation in M would alter the score on all these categories, and would spuriously raise the correlation unless these linked categories were concentrated in the same half of the test. (2) They derived separate sets of weights from the comparison of Normals vs. Schizophrenics, Nurses vs. Schizophrenics, and other groups. The correlation between the scoring weights is high, which they take as evidence for reliability (pp. 112 ff.). At least one serious objection is that the weights were derived in part from the same cases. If, by sampling alone, FK happened to be rare among the Schizophrenic group, this would cause the sign FK to have a weight in both the Normal-Schizophrenic key and the Nurse-Schizophrenic key. The evidence is not adequate to show that the weights would be the same if the two keys were

independently derived. This objection does not apply to another comparison of the same general type, where the four samples involved had no overlap. (3) Certain papers were scored repeatedly, using sets of weights derived in comparable but slightly different ways (p. 116). The correlations of the resulting sets of scores are advanced as evidence of reliability. Any correlation of separate scorings of the same set of responses is in part spurious. If responses of individual subjects were determined solely by chance, there would still be a correlation when keys having any similarity to each other were applied to the papers. The reliability of the performance of the subject, and that is what reliability coefficients are supposed to report, cannot be revealed by rescorings of the same performance.

CONCLUSIONS

The foregoing analysis and the appended bibliography are convincing evidence that Rorschach workers have sought statistical confirmation for their hypotheses. But the analysis also shows that the studies have been open to errors of two types: (1) erroneous procedures have led to claims of significance and interpretations which were unwarranted; and (2) failure to apply the most incisive statistical tests has led workers to reject significant relationships. So widespread are errors and unhappy choices of statistical procedures that few of the conclusions from statistical studies of the Rorschach test can be trusted. A few workers have been consistently sound in their statistical approach. But some of the most extensive studies and some of the most widely cited are riddled with fallacy. If these studies are to form part of the base for psychological science, the data must be reinterpreted. Perhaps ninety per cent of the conclusions so far published as a result of statistical Rorschach studies are unsubstantiated—not necessarily false, but based on unsound analysis.

Few of the errors were obvious violations of statistical rules. The Rorschach test is unlike conventional instruments and introduces problems not ordinarily encountered. Moreover, statistical methods for such tests have not been fully developed (11). It is most important that research workers using the Rorschach secure the best possible statistical guidance, and that editors and readers scrutinize studies of the test with great care. But statisticians have a responsibility too, to examine the logic of Rorschach research and the peculiar character of clinical tests, in order to sense the limitations of conventional and mathematically sound procedures.

Present statistical tools are imperfect. And no procedure is equally advisable for all studies. Within these limitations, this review has suggested the following guides to future practice.

1. Matching procedures in which a clinical synthesis of each Rorschach record is compared with a criterion are especially appropriate.

2. If ratings are to be treated statistically, it is often advisable to dichotomize the rating and apply chi-square or bi-serial r .
3. Common errors which must be avoided in significance tests are:
 - a. Use of critical ratio and uncorrected chi-square for unsuitably small samples.
 - b. Use of sample values in the formula for differences between proportions.
 - c. Use of formulas for independent samples when matched samples are compared.
 - d. Interpretation of P -values without regard for the inflation of probabilities when hundreds of significance tests are made or implicitly discarded.
 - e. Acceptance of conclusions when a significant difference is found with a hypothesis based on fluctuations in a particular sample.
4. Counting procedures are in general preferable to additive methods for Rorschach data. The most widely useful procedures are chi-square and analysis of differences in mean rank. These yield results which are invariant when scores are transformed.
5. Normalizing scores is frequently desirable before making significance tests involving variance.
6. Where groups differ in total number of responses, this factor must be held constant before other differences can be soundly interpreted. Three devices for doing this are: rescoring a fixed number of responses on all papers, constructing subgroups equated on the number of responses, and analyzing profiles of normalized scores (pattern tabulation).
7. Ratio and difference scores should rarely be used as a basis for statistical analysis. Instead, patterns should be defined and statistical comparisons made of the frequency of a certain pattern in each group. Use of chi-square with frequencies of Rorschach "signs" is recommended.
8. Multiple regression and linear discriminant functions are unlikely to reveal the relationships of Rorschach scores with other variables, since the assumption of linear compensation is contrary to the test theory.
9. Rank correlation, curvilinear correlation, or correlation of normalized scores are often more suitable than product-moment correlation.
10. No entirely suitable method for estimating Rorschach reliability now exists. Studies in this area are much needed.

There are in the Rorschach literature numerous encouraging bits of evidence. The question whether the test has any merit seems adequately answered in the affirmative by studies like those of Troup, Judith Krugman, Williams (69), and Munroe. Supplemented as these are by the testimony of intelligent clinical users of the test, there is every reason to treat the test with respect. One cannot attack the test merely because most Rorschach hypotheses are still in a pre-research stage. Some of the studies which failed to find relationships might have supported Rorschach theory if the analysis had been more perfect. How accurate the test is, how particular combinations of scores are to be interpreted, and how to use Rorschach data in making predictions about groups are problems worth considerable effort. With improve-

ments in projective tests, in personality theory, and in the statistical procedures for verifying that theory, we can look forward to impressive dividends.

BIBLIOGRAPHY

1. ABEL, T. M. Group Rorschach testing in a vocational high school. *Rorschach Res. Exch.*, 1945, 9, 178-188.
2. BECK, S. J. Personality structure in schizophrenia. *Nerv. and ment. Dis. Monogr.*, 1938, No. 63.
3. BROWER, D. The relation between certain Rorschach factors and cardiovascular activity before and after visuo-motor conflict. *J. gen. Psychol.*, 1947, 37, 93-95.
4. BROWN, R. R. The effect of morphine upon the Rorschach pattern in post-addicts. *Amer. J. Orthopsychiat.*, 1943, 13, 339-342.
5. BUHLER, C., BUHLER, K., & LEFEVER, D. W. *Rorschach standardization studies. Number I. Development of the Basic Rorschach Score.* Los Angeles: C. Buhler, 1948.
6. COCHRAN, W. G. The chi-square correction for continuity. *Iowa St. Col. J. Sci.*, 1942, 16, 421-436.
7. CRONBACH, L. J. The reliability of ratio scores. *Educ. psychol. Msmt.*, 1941, 1, 269-278.
8. CRONBACH, L. J. Note on the reliability of ratio scores. *Educ. psychol. Msmt.*, 1943, 3, 67-70.
9. CRONBACH, L. J. A validation design for personality study. *J. consult. Psychol.*, 1948, 12, 365-374.
10. CRONBACH, L. J. Pattern tabulation: a statistical method for treatment of limited patterns of scores, with particular reference to the Rorschach test. *Educ. psychol. Msmt.*, in press.
11. CRONBACH, L. J. Statistical methods for multi-score tests. Paper presented before the Biometrics Section, American Statistical Association, December, 1948. To be published.
12. DAVIDSON, HELEN H. *Personality and economic background.* New York: King's Crown Press, 1945.
13. EDWARDS, A. L. Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. *Psychometrika*, 1948, 13, 185-187.
14. FESTINGER, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, 11, 97-105.
15. FOSBERG, I. A. Rorschach reactions under varied instructions. *Rorschach Res. Exch.*, 1938, 3, 12-31.
16. FOSBERG, I. A. An experimental study of the reliability of the Rorschach technique. *Rorschach Res. Exch.*, 1941, 5, 72-84.
17. FREEMAN, H., RODNICK, E. H., SHAKOW, D., & LEBEAUX, T. The carbohydrate tolerance of mentally disturbed soldiers. *Psychosom. Med.*, 1944, 6, 311-317.
18. GANN, E. *Reading difficulty and personality organization.* New York: King's Crown Press, 1945.
19. GOLDFARB, W. A. A definition and validation of obsessional trends in the Rorschach examination of adolescents. *Rorschach Res. Exch.*, 1943, 7, 81-108.
20. GOLDFARB, W. Effects of early institutional care on adolescent personality. *Amer. J. Orthopsychiat.*, 1944, 14, 441-447.
21. GUILFORD, J. P. (Ed.) *Printed classification tests.* AAF Aviation Psychology Program Research Reports, No. 3. Washington: Government Printing Office, 1947.
22. GUSTAV, ALICE. Estimation of Rorschach scoring categories by means

- of an objective inventory. *J. Psychol.*, 1946, **22**, 253-260.
23. HARRIS, R. E., & CHRISTIANSEN, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269-284.
 24. HARRIS, T. M. The use of projective techniques in industrial selection. In *Exploring individual differences*, American Council on Education Studies, Series 1, No. 32, 1948. Pp. 43-51.
 25. HERTZ, MARGUERITE R. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: I. The movement factors. *J. gen. Psychol.*, 1942, **27**, 119-188.
 26. HERTZMAN, M. A comparison of the individual and group Rorschach tests. *Rorschach Res. Exch.*, 1942, **6**, 89-108.
 27. HERTZMAN, M., & MARGULIES, H. Developmental changes as reflected in Rorschach test responses. *J. genet. Psychol.*, 1943, **62**, 189-215.
 28. HERTZMAN, M., ORLANSKY, J., & SEITZ, C. P. Personality organization and anoxia tolerance. *Psychosom. Med.*, 1944, **6**, 317-331.
 29. KABACK, G. R. *Vocational personalities: an application of the Rorschach group method*. New York: Bureau of Publications, Teachers Coll., Columbia Univ., 1946.
 30. KELLEY, D. M., MARGULIES, H., & BARRERA, S. E. The stability of the Rorschach method as demonstrated in electric convulsive therapy cases. *Rorschach Res. Exch.*, **5**, 1941, 35-43.
 31. KRUGMAN, J. I. A clinical validation of the Rorschach with problem children. *Rorschach Res. Exch.*, 1942, **6**, 61-70.
 32. KRUGMAN, M. Psychosomatic study of fifty stuttering children. *Amer. J. Orthopsychiat.*, 1946, **16**, 127-133.
 33. KURTZ, A. K. A research test of the Rorschach test. *Personnel Psychol.*, 1948, **1**, 41-51.
 34. LEVERETT, H. M. Table of mean deviates for various portions of the unit normal distribution. *Psychometrika*, 1947, **12**, 141-152.
 35. LINDQUIST, E. F. *A first course in statistics*. (Revised Ed.) Boston: Houghton Mifflin, 1942.
 36. McCANDLESS, B. R. The Rorschach as a predictor of academic success. *J. appl. Psychol.*, 1949, **33**, 43-50.
 37. McNEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, **12**, 153-157.
 38. MARGULIES, H. Rorschach responses of successful and unsuccessful students. *Arch. Psychol.*, N. Y., No. 271. New York, 1942.
 39. MELTZER, H. Personality differences between stuttering and non-stuttering children. *J. Psychol.*, 1944, **17**, 39-59.
 40. MONTALTO, F. D. An application of the Group Rorschach technique to the problem of achievement in college. *J. clin. Psychol.*, 1946, **2**, 254-260.
 41. MUNROE, RUTH L. Objective methods and the Rorschach blots. *Rorschach Res. Exch.*, 1945, **9**, 59-73.
 42. MUNROE, RUTH L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, 1945, No. 7.
 43. MUNROE, RUTH L. Rorschach findings on college students showing different constellations of subscores on the A.C.E. *J. consult. Psychol.*, 1946, **10**, 301-316.
 44. PEATMAN, J. G. *Descriptive and sampling statistics*. New York: Harper, 1947.
 45. PENROSE, L. S. Some notes on discrimination. *Ann. Eugenics*, 1947, **13**, 228-237.
 46. PIOTROWSKI, Z., CANDEE, B., BALINSKY, B., HOLTZBERG, S., & VON ARNOLD, B. Rorschach signs in the

- selection of outstanding young male mechanical workers. *J. Psychol.*, 1944, 18, 131-150.
47. RAPAPORT, D. *Diagnostic psychological testing, Vol. II*. Chicago: Year Book Publishers, 1946.
48. RICHARDSON, L. H. The personality of stutterers. *Psychol. Monogr.*, 1944, 56, No. 7.
49. RICKERS-OVSIANKINA, M. The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, 17, 227-257.
50. ROSS, W. D. The contribution of the Rorschach method to clinical diagnosis. *J. ment. Sci.*, 1941, 87, 331-348.
51. ROSS, W. D., FERGUSON, G. A., & CHALKE, F. C. R. The Group Rorschach Test in officer selection. *Bull. Canad. Psychol. Assn.*, 1945, 84-86.
52. ROSS, W. D., & ROSS, S. Some Rorschach ratings of clinical value. *Rorschach Res. Exch.*, 1944, 8, 1-9.
53. SARBIN, T. R., & MADOW, L. W. Predicting the depth of hypnosis by means of the Rorschach test. *Amer. J. Orthopsychiat.*, 1942, 12, 268-271.
54. SCHMIDT, H. O. Test profiles as a diagnostic aid: the Rorschach. *J. clin. Psychol.*, 1945, 1, 222-227.
55. SIEGEL, M. G. The diagnostic and prognostic validity of the Rorschach test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, 18, 119-133.
56. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State College Press, 1940.
57. SWIFT, J. W. Reliability of Rorschach scoring categories with preschool children. *Child Developm.*, 1944, 15, 207-216.
58. SWIFT, J. W. Rorschach responses of 82 pre-school children. *Rorschach Res. Exch.*, 1945, 9, 74-84.
59. SWINEFORD, F. A table for estimating the significance of the difference between correlated percentages. *Psychometrika*, 1948, 13, 23-25.
60. THOMPSON, G. M. College grades and the Group Rorschach. *J. appl. Psychol.*, 1948, 32, 398-407.
61. THORNTON, G. R., & GUILFORD, J. P. The reliability and meaning of Erlebnistypus scores on the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, 31, 324-330.
62. TROUP, E. A comparative study by means of the Rorschach method of personality development in twenty pairs of identical twins. *Genet. psychol. Monogr.*, 1938, 20, 461-556.
63. TULCHIN, S., & LEVY, D. Rorschach test differences in a group of Spanish and English refugee children. *Amer. J. Orthopsychiat.*, 1945, 15, 361-368.
64. VAUGHN, J., & KRUG, O. The analytic character of the Rorschach inkblot test. *Amer. J. Orthopsychiat.*, 1938, 8, 220-229.
65. VERNON, P. E. The Rorschach inkblot test. *Brit. J. med. Psychol.*, 1933, 13, 179-200.
66. VERNON, P. E. The matching method applied to investigations of personality. *Psychol. Bull.*, 1936, 33, 149-177.
67. WALKER, HELEN M. *Elementary statistical methods*. New York: Holt, 1943.
68. WERNER, H. Perceptual behavior of brain-injured, mentally defective children. *Genet. Psychol. Monogr.*, 1945, 31, 51-110.
69. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, 11, 21-29.
70. YATES, F. The analysis of contingency tables with groupings based on quantitative characters. *Biometrika*, 1948, 35, 176-181.

BOOK REVIEWS

Hsu, Francis, L. K. *Under the ancestors' shadow: Chinese culture and personality*. New York: Columbia University Press, 1948. Pp. xiv+317. \$3.75

Under the Ancestors' Shadow is an anthropological (or sociological) study of "West Town," a small semi-rural community in southwest China. The author expresses his conviction that "the essential social structure" of West Town is typical of China as a whole, although he admits that this judgment can only be confirmed through further study.

The book is appropriately titled: Hsu's account of West Town culture is in some respects far from adequate (in the judgment of a social psychologist), but he does a highly effective job of giving substance and credibility to his central theme as expressed in the title. Taking up one after another aspect of the culture, he elaborates the picture of a people who live "under the ancestors' shadow," and traces the culture patterns which flow from this way of life: the cult of ancestors, family unity, submission to authority, father-son identification, filial piety, and so on.

The economic bases of life in West Town are inexpertly treated. Processes of cultural change, though mentioned a number of times in passing, are dealt with in perfunctory fashion. But for readers of psychological background the chief inadequacies of the book stem from the fact that the author is first and foremost a careful recorder of what Linton has called the *overt* aspects of the culture, and is not particularly skillful in getting at the psychological variables which underlie these overt aspects. There are two shortcomings involved: first, he does not seem skilled at gathering the kinds of data which would permit inferences concerning the psychological variables; and second, when he has such data available, he does not always interpret them with a sure hand. The fact that the book is sub-titled "Chinese Culture and Personality" leads the reader to expect a study in the manner of Linton, Kluckhohn, or Dubois, rich in materials of interest to the psychologist. It falls short of the mark. In dealing with personality, Hsu writes like a man who has been handed his conceptual tools after he had completed his field study.

There is a good deal of internal evidence to indicate that Hsu is a conscientious observer and recorder; but on the whole the study cannot be commended on methodological grounds. Hsu fails to employ the various simple forms of controlled or systematic observation which would have been readily available to him. His treatment of personality structure would have gained immensely from the inclusion of a few autobiographies. He offers little or no information as to his informants. (The reviewer can see no reason why such ethnological field reports as this should not include as a matter of course an "essay on informants," no matter how unpretentious: What sorts of data were obtained from direct observation and what from informants? What kinds of persons

supplied what kinds of information? What kinds of persons proved the best and most reliable sources of information? What sorts of resistance and detectable falsifying appeared in informant reports?)

In spite of these shortcomings, *Under the Ancestors' Shadow* is a useful addition to the small but growing shelf of field reports on Chinese culture.

JOHN W. GARDNER.

Carnegie Corporation of New York.

BOOKS AND MATERIALS RECEIVED

ANDERSON, JOHN E. *The psychology of development and personal adjustment*. New York: Henry Holt, 1949. Pp. xvi+720. \$3.25.

CAVAN, RUTH, et al. *Personal adjustment in old age*. Chicago: Science Research Associates, 1949. Pp. xiii+204. \$2.95.

DENNIS, WAYNE. *Readings in general psychology*. New York: Prentice-Hall, 1949. Pp. xi+525. \$3.75.

DONAHUE, WILMA T., et al. *The measurement of student adjustment and achievement*. Ann Arbor: Univ. of Michigan Press, 1949. Pp. 256. \$3.75 cloth. \$3.00 paper.

FETTERMAN, JOSEPH L. *Practical lessons in psychiatry*. Springfield, Ill.: Charles C Thomas, 1949. Pp. ix+342. \$5.75.

HOCH, PAUL, AND ZUBIN, JOSEPH. *Psychological development in health and disease*. New York: Grune & Stratton, 1949. Pp. viii+283. \$4.50.

HOVLAND, CARL I., LUMSDAINE, ARTHUR A., AND SHEFFIELD, FRED D. *The American Soldier, Experiments on mass communication*. Vol. III. Princeton: Princeton Univ. Press, 1949. Pp. x+345. \$5.00.

HUMPHREY, GEORGE, *Directed thinking*. New York: Dodd, Mead, 1949. Pp. 229. \$3.50.

JENKINS, G. G., SHACTER, H., AND BAUER, W. W. *These are our children*. Chicago: Scott, Foresman, 1949. Pp. 192 (200 photographs). \$3.50.

JOHNSON, PALMER O. *Statistical methods in research*. New York: Prentice-Hall, 1949. Pp. xvi+377.

JONES, HAROLD E. *Motor performance and growth*. Berkeley: Univ. of California Press, 1949. Pp. x+181. \$3.00.

KANNER, LEO. *A miniature textbook of feeble-mindedness*. Child Care Monographs No. 1. New York: Child Care Publications, 1949. Pp. 33. \$1.75.

LAPIERE, R. T., AND FARNSWORTH, P. R. *Social psychology*. (3rd Ed.) New York: McGraw-Hill, 1949. Pp. xiii+626. \$4.50.

LINDESMITH, A. R., AND STRAUSS, A. L. *Social psychology*. New York: Dryden Press, 1949. Pp. xvi+549. \$4.50.

LIPPITT, RONALD. *Training in community relations*. New York: Harper, 1949. Pp. xiv+286. \$3.50.

MCKINNEY, FRED. *The psychology of personal adjustment*. (2nd Ed.) New York: John Wiley, 1949. Pp. xi+752. \$6.00.

MATHEWSON, ROBERT H. *Guidance policy and practice*. New York: Harper, 1949. Pp. xiv+294. \$3.00.

MOON, G. R. *How to become a doctor*. Philadelphia: Blakiston, 1949. Pp. ix+131. \$2.00.

MURSELL, JAMES L. *Psychological testing*. (2nd Ed.) New York: Longmans, Green, 1949. Pp. xvi+473. \$4.00.

PASTORE, NICHOLAS. *The nature-nurture controversy*. New York: Columbia Univ. Press, 1949. Pp. xvi+213. \$3.25.

SANDERS, D. C. *Student selection and academic success*. Sidney, Australia: T. H. Tennant, 1948. Pp. 158.

STENDLER, CELIA B. *Children of Brasstown*. Urbana: Univ. of Illinois Press, 1941. Pp. 103.

SUPER, DONALD E. *Appraising vocational fitness*. New York: Harper, 1949. Pp. xxi+727.

UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949. Pp. vii+638. \$4.50.

WALKER, K., AND STRAUSS, E. B. *Sexual disorders in the male*. (3rd Ed.) Baltimore: Williams & Wilkins, 1948. Pp. xiii+269. \$3.50.

WARNER, L. W., MEEKER, M., AND EELLS, K. *Social class in America*. Chicago: Social Science Research Council, 1949. Pp. xiii+274. \$4.25.

